Análise forense e segurança facial baseada em dados

Combatendo a falsificação: para detectar deepfake com sinais biológicos multidimensionais

Xin Lei Jin, 1 **Dengpan Ye**, 1 e Chuanxi Chen1

Mostre mais

Editor Acadêmico: Pequim Chen

Recebido27 de dezembro de 2020

Revisado20 de março de 2021

Aceitaram10 de abril de 2021

Publicados22 de abril de 2021

Abstrato

A tecnologia deepfake é convenientemente abusada com o baixo limiar tecnológico, o que pode trazer enormes riscos para a segurança social. À medida que a tecnologia de síntese baseada em GAN se torna mais forte, vários métodos são difíceis de classificar eficazmente o conteúdo falso. No entanto, embora o conteúdo falso gerado pelos GANs possa enganar os olhos humanos, ele ignora os sinais biológicos ocultos no vídeo facial. Neste artigo, propusemos um novo método de vídeo forense com sinais biológicos multidimensionais, que extrai a diferença do sinal biológico entre vídeos reais e falsos a partir de três dimensões. Os resultados experimentais mostram que nosso método atinge 98% de precisão no principal conjunto de dados públicos. Comparado com outras tecnologias, o método proposto extrai apenas informações de vídeo falsas e não se limita a um método de geração específico,

1. Introdução

Com o rápido avanço da visão computacional e da tecnologia de processamento de conteúdo digital, a adulteração de rostos não se limita mais a imagens, algumas tecnologias de aprendizagem profunda (por exemplo, deepfake) podem ser

utilizadas para gerar rostos humanos em vídeos, que são muito semelhantes aos vídeos de rostos naturais feitos usando câmeras digitais, mas é difícil distingui-las a olho nu. O recente estudo de Korshunov [1] mostra que vídeos falsos podem facilmente enganar o sistema de reconhecimento facial, e alguns riscos graves de segurança, como notícias falsas, foram levantados por eles.

A tecnologia Deepfake é o resultado do progresso científico e tecnológico e do rápido desenvolvimento da tecnologia de inteligência artificial, e tem amplas perspectivas de aplicação. Por exemplo, a tecnologia deepfake é usada em indústrias de entretenimento, como filmes, o que pode economizar tempo e custos de mão de obra. No entanto, se esta tecnologia for abusada por criminosos, também causará uma grave crise, podendo até forjar os discursos dos líderes mundiais, colocando seriamente em perigo a segurança política. Portanto, a análise forense de vídeos deepfake é de grande importância. Atualmente, o método forense de vídeo deepfake é baseado principalmente em informações intraframe ou interframe, analisando a diferença entre vídeos reais e falsos.

Neste artigo, propomos um método forense de vídeo deepfake baseado em sinais biológicos multidimensionais. Trabalhos recentes mostram que os sinais de frequência cardíaca podem ser usados para distinguir com eficácia entre vídeos reais e falsos [2 , 3]. Embora os GANs possam gerar conteúdo falso que engana os olhos humanos, eles destroem os sinais biológicos originais do vídeo real, como os sinais de frequência cardíaca. Portanto, podemos classificar os vídeos reais e falsos extraindo e analisando os sinais biológicos dos vídeos. Nossas principais contribuições são as seguintes:(1)Propomos um método forense de vídeo sintético, que analisa principalmente os diferentes sinais biológicos entre vídeos reais e falsos para detectar o conteúdo falsificado.(2)Exploramos ainda mais as informações distintas no cenário multidimensional para garantir a eficiência tecnológica. Ou seja, utilizamos o espaço RGB para nos concentrarmos nas variações de cores, o espaço YUV para nos concentrarmos na alteração do brilho e o método de crominância para reduzir os efeitos de ruído.(3)Analisamos as deficiências da fotopletismografia tradicional (PPG) e usamos uma rede neural profunda para realizar a classificação de vídeos reais e falsos. Os resultados experimentais mostram que os modelos profundos podem atingir alta precisão de detecção, que é de cerca de 98% no principal conjunto de dados públicos.

O restante deste trabalho está organizado da seguinte forma. A seção 2 apresenta trabalhos relacionados, incluindo o desenvolvimento de PPG e análise forense de vídeo deepfake. A seção 3 descreve detalhadamente o método proposto. A

seção 4 mostra os detalhes e resultados de nosso experimento. Na Seção 5 , concluímos e apresentamos o trabalho futuro.

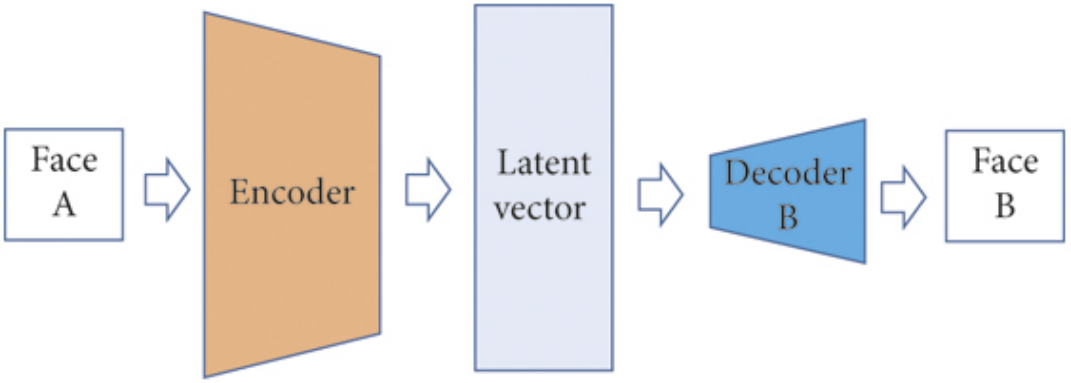
2. Trabalho relacionado

2.1. Profundo falso

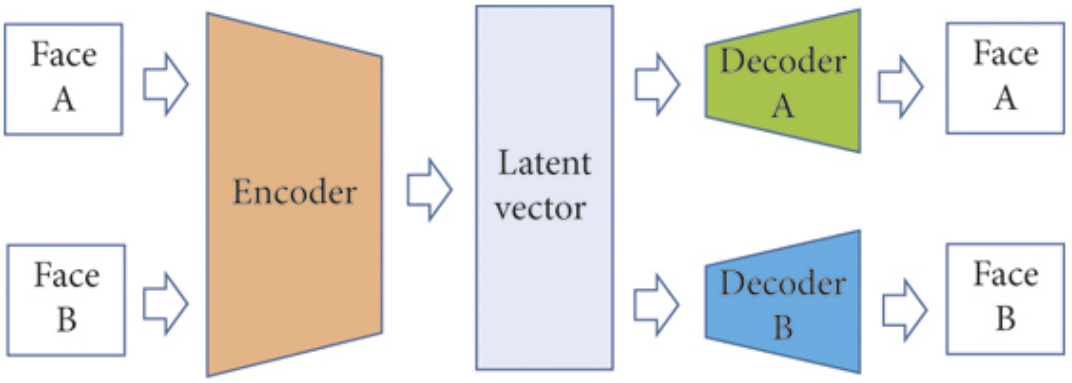
Deepfakes são vídeos falsos manipulados digitalmente para retratar pessoas dizendo e fazendo coisas que nunca aconteceram de fato. Deepfakes dependem de redes neurais que analisam grandes conjuntos de amostras de dados para aprender a imitar as expressões faciais e maneirismos de uma pessoa. O processo envolve alimentar imagens de duas pessoas em um algoritmo de aprendizado profundo para treiná-las para trocar rostos.

O pipeline geral do deepfake básico é mostrado na Figura 1. O autoencoder geralmente é formado por duas redes neurais convolucionais (o codificador e o decodificador). O codificador converte a face do alvo de entrada em um vetor. Existe apenas um único codificador, independentemente das identidades dos sujeitos, para garantir que o codificador capture atributos independentes da identidade, como expressões faciais. Por outro lado, cada identidade possui um decodificador dedicado, que gera uma face do sujeito correspondente a partir do vetor. Especificamente, um par codificador-decodificador é formado alternativamente usando codificador e decodificador para face de entrada de cada sujeito, e seus parâmetros são otimizados para minimizar os erros de reconstrução. A atualização dos parâmetros é realizada com retropropagação até a convergência. A fase de treinamento pode ser definida comoonde L denota o valor de perda do autoencoder; N é o número de dados de entrada da rede; $F\ i$ é a imagem da face de entrada; θ é o peso do codificador E ; e Φ são os pesos do decodificador D .

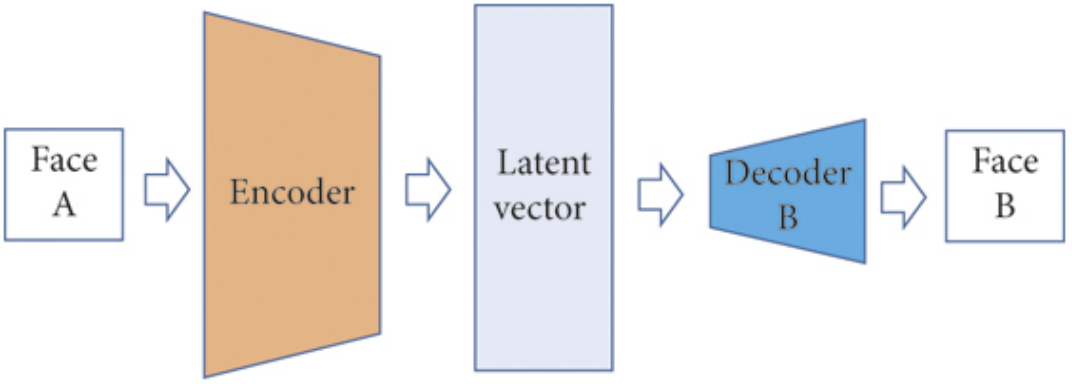
(b)



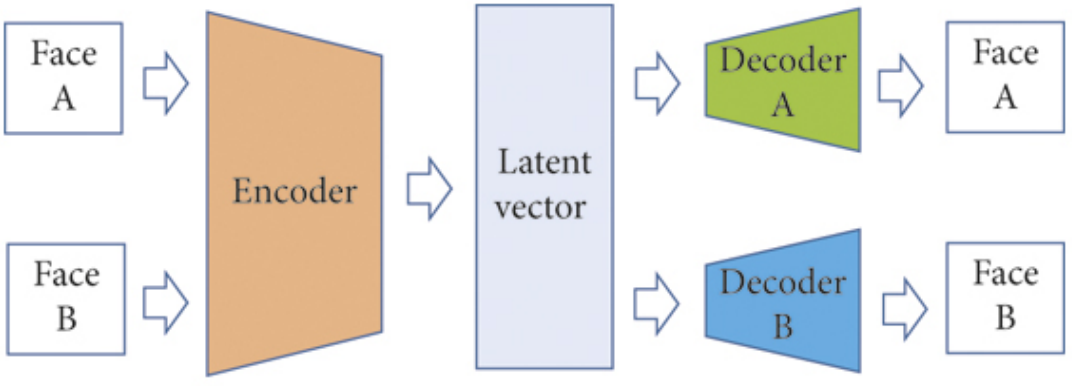
(a)


Face A
Face B
Encoder
Latent vector
Decoder A
Decoder B
Face A
Face B

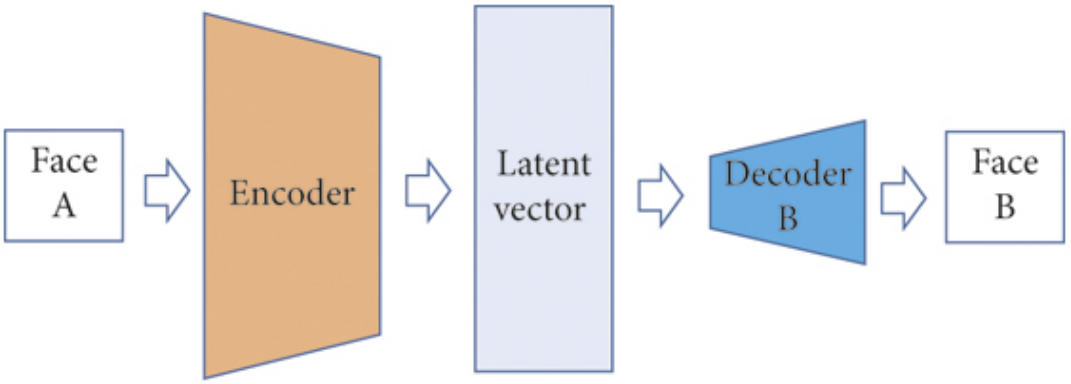
(b)


Face A
Encoder
Latent vector
Decoder B
Face B

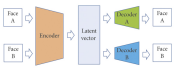
(a)


Face A
Face B
Encoder
Latent vector
Decoder A
Decoder B
Face A
Face B

(b)



- **(a)**



- **(b)**

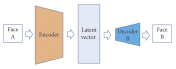


figura 1

Visão geral do procedimento deepfake. (a) A fase de treinamento do deepfake. (b) O estágio de conversão do deepfake.

No estágio de conversão, o decodificador treinado B é usado para decodificar o vetor latente da face A para obter a imagem de troca de face de A. Da mesma forma, podemos usar o decodificador treinado A para decodificar o vetor latente da face B para obter a face -troca de imagem de B. O estágio de conversão pode ser declarado comoonde *F* denota a face original e *F ′* denota a face falsa.

2.2. Sinais Biológicos

A extração de sinal biológico foi originalmente usada na área médica para detectar se a frequência cardíaca (FC) do paciente ou outros sinais são normais, para que o médico possa observar o sinal biológico anormal do paciente a tempo. No entanto, derivações de eletrocardiograma (ECG), oxímetros de pulso e outros detectores exigem que sensores específicos sejam conectados ao corpo humano. Para evitar o uso de sensores intrusivos, pesquisadores de visão computacional propuseram um método de medições remotas de FC sem contato, baseado na observação de

mudanças sutis na cor e no movimento no vídeo RGB, como a fotopletismografia remota (PPG) [4 , 5] .

Balakrishnan et al. [6] mostram que a atividade cardíaca pode causar movimentos da cabeça, que podem ser usados para extrair estimativas de frequência cardíaca de fluxos de vídeo. Tulyakov propôs um método de crominância, que pode efetivamente melhorar a precisão da estimativa da frequência cardíaca [5]. Niu propôs um método de estimativa remota da frequência cardíaca baseado em aprendizado profundo e obteve bons resultados [7].

2.3. Detecção de falsificação

Para lidar com os possíveis danos causados pelos vídeos deepfake, os pesquisadores estão explorando métodos eficazes para classificar vídeos reais e falsos. Como o deepfake também é uma falsificação de imagens, os métodos de detecção precoce podem aprender com o método de detecção de falsificações de imagens. Recentemente, vários detectores de alta eficiência com os novos algoritmos foram propostos para melhorar o desempenho da detecção e localização de adulteração [8 , 9]. Além disso, para detectar especificamente falsificações de deepfake, os pesquisadores classificam vídeos reais e falsos com base em informações intraframe, informações interframe ou artefatos especiais.

Nguyen et al. [10] propuseram uma rede cápsula que pode detectar vários tipos de ataques, desde ataques de apresentação usando imagens impressas e vídeos reproduzidos até ataques usando vídeos falsos criados usando aprendizado profundo. Ela usa menos parâmetros do que as redes neurais convolucionais tradicionais com desempenho semelhante. Faça et al. [11] usaram uma rede neural convolucional profunda (VGGFace) para detectar imagens reais/falsas de GANs. Afchar et al. [12] exploraram recursos em nível mesoscópico, em vez de recursos puramente microscópicos e macroscópicos, e propuseram mesoneto e rede meso-4, que possuem um baixo número de parâmetros. Bonetini et al. [13] combinaram CNNs, camadas de atenção e treinamento siamês e obtiveram bom desempenho no DFDC. Li e Lyu [14] criou dados negativos apenas usando uma operação simples de processamento de imagem, em vez de usar deepfake para produzir, e então usou modelos CNN para classificar os vídeos. Zhao [15] formulou a detecção de deepfake como um problema de classificação refinado e propôs uma nova rede multiatencional de detecção de deepfake. Liu [16] propôs um novo método Spatial-Phase Shallow Learning (SPSL), que combina a imagem espacial e o espectro de fase para capturar os artefatos de upsampling da falsificação de rosto para melhorar a transferibilidade.

Güera e Delp [17], baseados em inconsistências temporais entre quadros, utilizaram CNN (extração de características de quadros) e RNN (análise de sequência temporal) para classificação de vídeos reais e falsos. Sabir et al. [18] também propuseram o método CNN + RNN, mas usaram alinhamento facial e recorrência bidirecional.

Agarwal et al. [19] rastrearam movimentos faciais e da cabeça e, em seguida, extraíram a presença e força de unidades de ação específicas e classificaram vídeos reais e falsos pelo SVM. Li et al. [20] usaram CNN e RNN para detectar piscadas anormais em vídeos falsos. Yang et al. [21] classificaram vídeos reais e falsos pela inconsistência das poses da cabeça em 3D. Li et al. [22] detectou se a imagem de entrada pode ser decomposta na mistura de duas imagens de fontes diferentes. Wang et al. com base no monitoramento do comportamento dos neurônios para detectar rostos falsos sintetizados por IA [23].

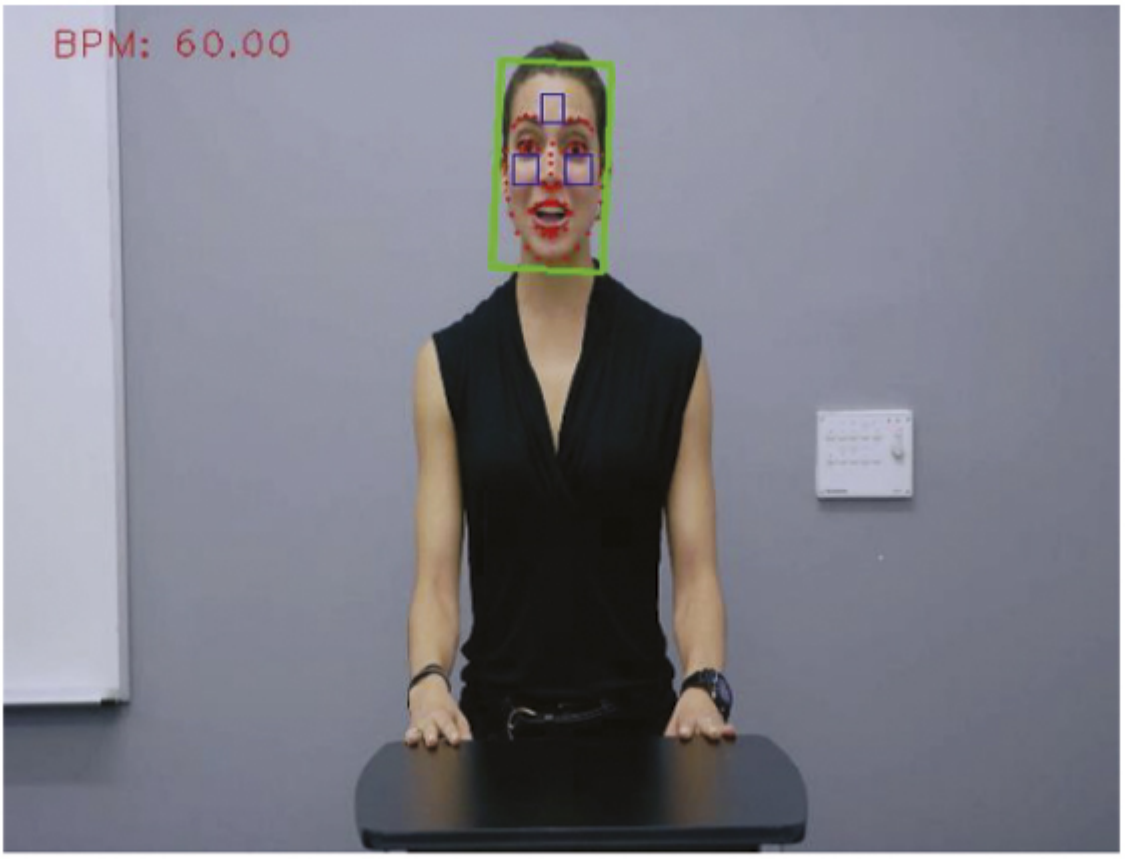
3. Método

Nesta seção, analisamos primeiro os sinais biológicos discrepantes entre vídeos reais e falsos. Em seguida, apontamos a ineficiência do método PPG tradicional para detecção de vídeo deepfake. Por fim, propomos um método forense de vídeo deepfake baseado na inconsistência de sinais biológicos, e os resultados experimentais da avaliação verificam a eficácia do nosso método.

3.1. Detecção de Deepfake com Sinais Biológicos

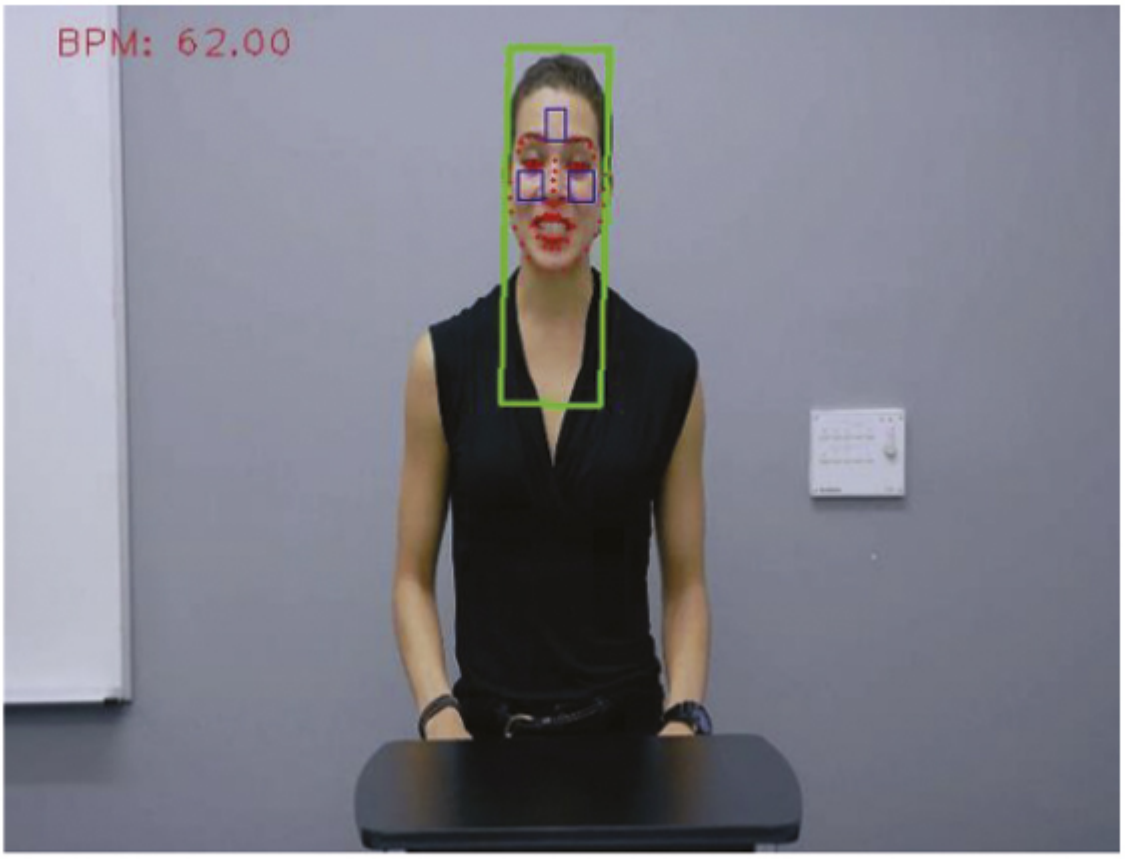
Embora a tecnologia PPG tenha sido desenvolvida há muito tempo, não é fácil extrair sinais de frequência cardíaca em um ambiente sem restrições. Analisamos o método de extração manual de sinais de frequência cardíaca do vídeo facial usando visão computacional; A Figura 2 mostra que esses métodos não conseguem distinguir vídeos falsos de vídeos reais. Selecionamos um par de vídeos reais e falsos do conjunto de dados DeepFakeDetection (DFD) e usamos o método do filtro Kalman [24] para estimar os sinais de frequência cardíaca deles. O resultado mostra que a diferença nos sinais de frequência cardíaca entre vídeos reais e falsos não é óbvia.

(c)



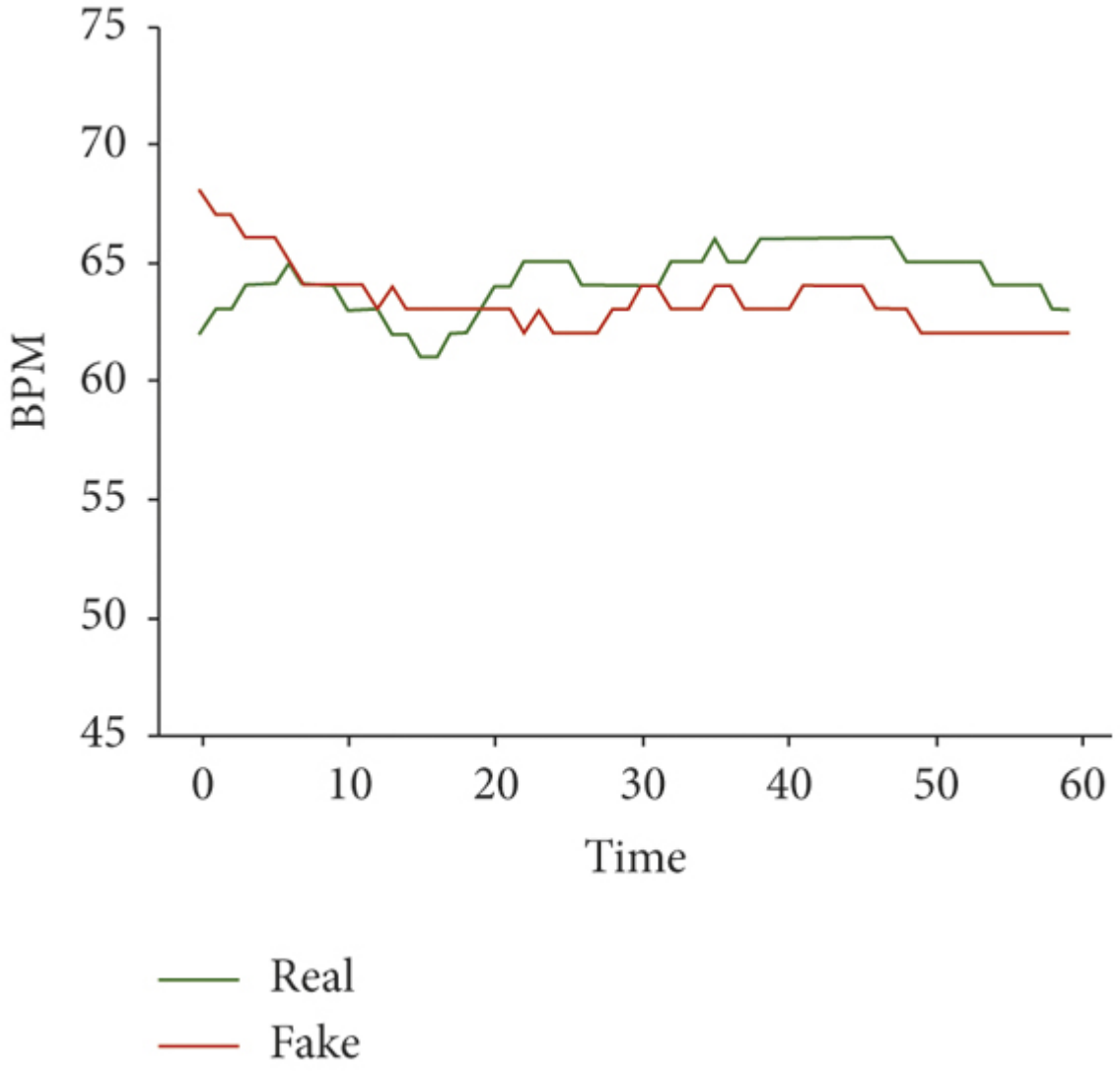
Fake video

(a)

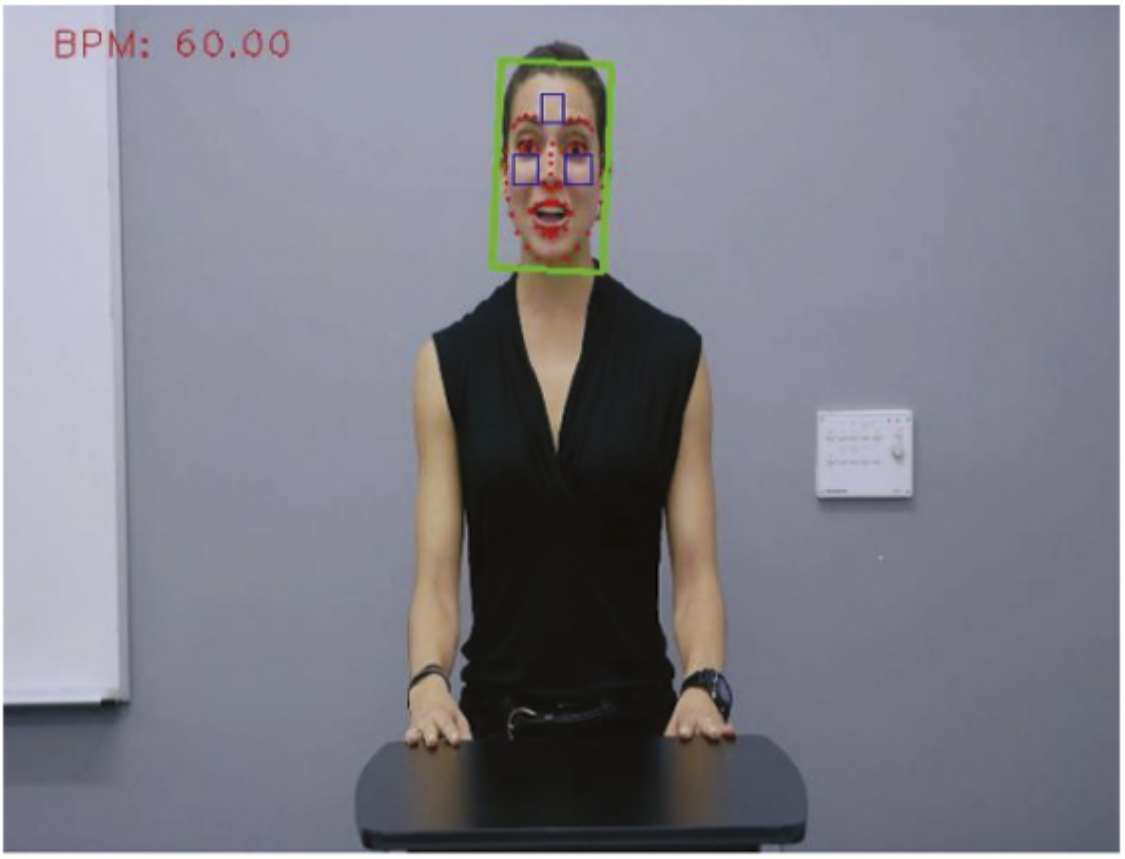


Real video

(b)

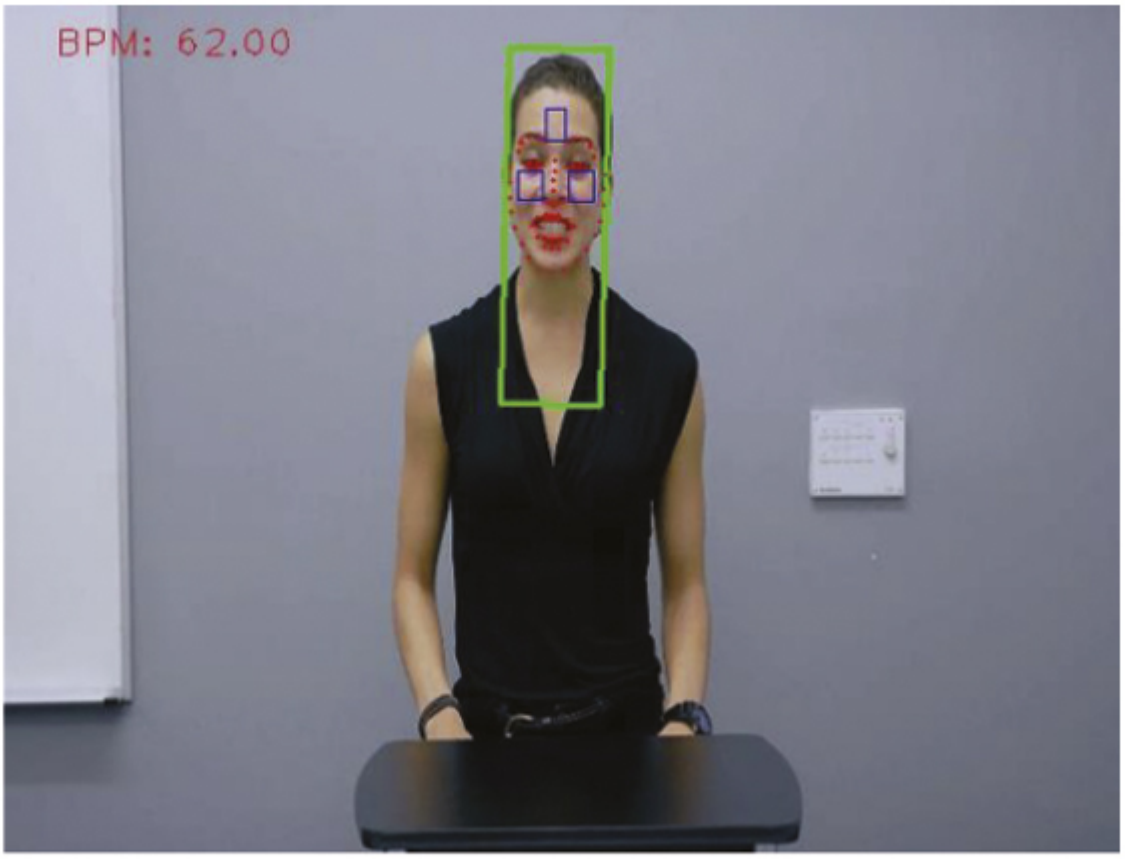


(c)



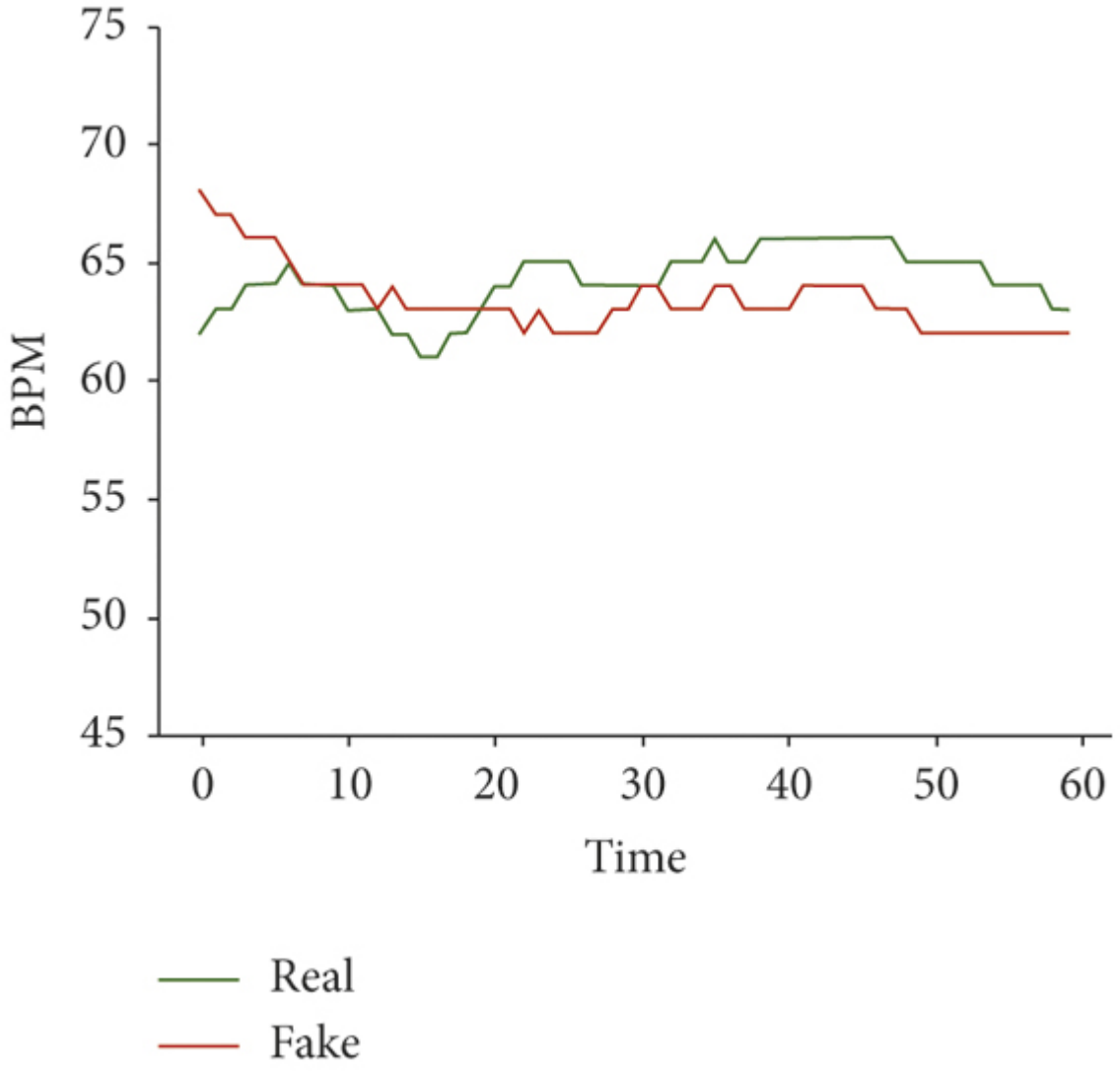
Fake video

(a)

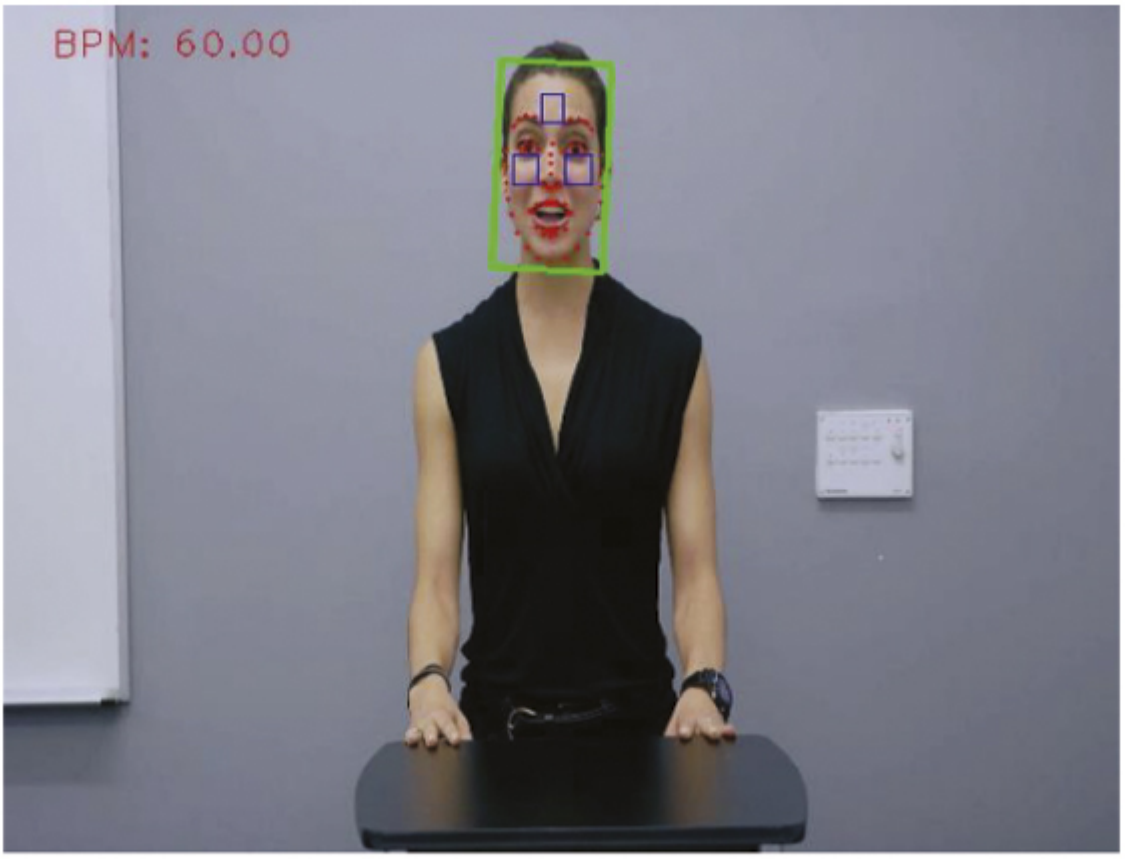


Real video

(b)



(c)

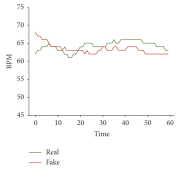


Fake video

- **(a)**

Real video

- **(b)**



- **(c)**

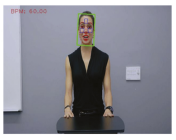


Figura 2

Comparação da frequência cardíaca de vídeo real e falsa. O eixo horizontal representa o número de quadros do vídeo e o eixo vertical representa a frequência cardíaca detectada.

Geralmente, para eliminar artefatos de movimento e ruídos causados por mudanças ambientais e extrair melhor os sinais puros de frequência cardíaca, os vídeos são sempre processados por remoção de ruído e filtragem. No entanto, essas tecnologias destroem os sinais anormais de frequência cardíaca no vídeo falso, o que causa o efeito de classificação fraco. Portanto, mapeamos o vídeo para ppg_map e o classificamos por meio da rede profunda para obter o efeito de classificação de vídeo deepfake com base em diferentes algoritmos de extração de frequência cardíaca. Em detalhes, dado um vídeo Vmmc5 (= {T1, T2, T3 …… Tk}) incluindo *k* quadros, para cada quadro, primeiro extraímos a face e fazemos o alinhamento da face. Em seguida, é realizada a segmentação da pele para retirar a influência do fundo. A seguir, a imagem do rosto é dividida em *n*blocos (R1, R2, R3… Rn), que são independentes entre si. Por último, calculamos o valor do sinal em cada bloco multidimensional. Os valores dos sinais de diferentes blocos no mesmo quadro são organizados em colunas, e os valores dos sinais do mesmo bloco em diferentes quadros são organizados em linhas para formar nosso ppg_map. Em seguida, esses ppg_maps são usados para treinar o modelo de classificação CNN, conforme mostrado na Figura 3 .

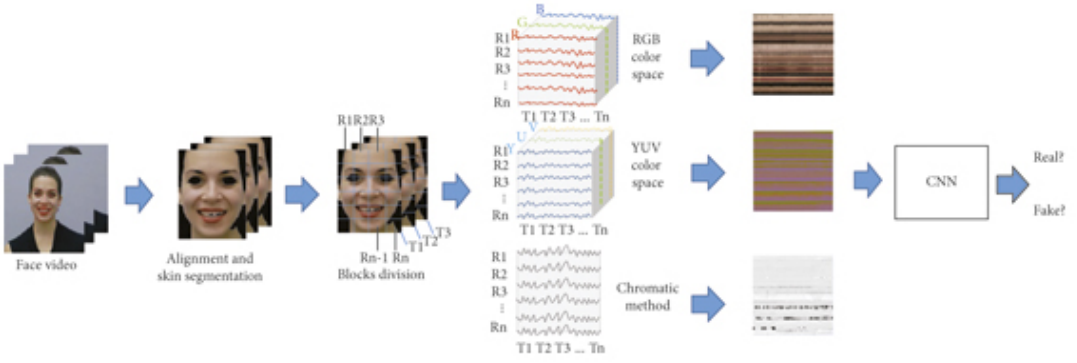


Figura 3

Visão geral do método proposto.

No processo de geração do ppg_map, é necessário evitar os efeitos adversos do movimento da cabeça e do fundo dos personagens. Discutiremos isso em detalhes na Seção 3.2 .

3.2. Geração ppg_Map

O batimento do coração humano causa a constrição periódica dos vasos sanguíneos, o que afeta o reflexo da luz na pele. Esta alteração não é facilmente detectável pelos olhos humanos, mas pode ser detectada e registrada por instrumentos ópticos. A área facial no vídeo facial pode refletir bem as informações da frequência cardíaca do corpo humano. Assim, localizamos a região facial e extraímos os sinais biológicos.

3.2.1. Detecção e alinhamento de rosto

Para tornar a detecção mais rápida e simples, o método Viola e Jones [25] é utilizado para detectar rostos humanos. No entanto, como os rostos no vídeo não serão fixados em uma determinada posição e ângulo, alinhamos os rostos detectados girando o rosto para manter os dois olhos no mesmo nível. Por outro lado, a área facial detectada pelo método Viola-Jones é maior que a área real da face e contém mais área de fundo; ajustamos ainda mais a região de interesse (ROI). Em outras palavras, localizamos 81 pontos de referência e usamos quatro pontos (1, 8, 15 e 71) como pontos de referência para ajustar a região da face (Figura 4) para fazer com que o ROI incluísse tantas regiões da face quanto possível.

Figura 4

Refinando o ROI facial. Localizamos 81 pontos de referência no rosto e usamos os quatro pontos (1, 8, 15 e 71) como pontos de referência para ajustar ainda mais a área facial detectada para reduzir a área de fundo.

3.2.2. Detecção e segmentação de pele

Os sinais biológicos são extraídos da pele facial, reduzindo assim a influência negativa de outras áreas não cutâneas, como olhos, cabelos e áreas de fundo. Entretanto, isto também reduzirá a perturbação causada pelo piscar dos olhos e pelos movimentos dos lábios. Consequentemente, no quadro do vídeo, primeiro adotamos o algoritmo de detecção de pele para obter as principais informações da pele facial. Em seguida, como máscara, a área da pele é usada para extrair a pele do rosto e remover o fundo e as áreas não cutâneas.

3.2.3. Divisão de Blocos e Extração de Sinais

Agora fizemos a detecção e segmentação da pele para deixar os sinais biológicos mais claros. Em seguida, o quadro de vídeo é dividido em m n blocos para extrair sinais biológicos de cada bloco. O método PPG extrai principalmente sinais de frequência cardíaca de três dimensões [4–6] . Ou seja, a dimensão RGB reflete intuitivamente as mudanças na cor do rosto humano, a dimensão YUV presta mais atenção às mudanças no brilho e a dimensão crominância pode efetivamente eliminar artefatos ambientais e erros causados pelo movimento da cabeça. Assim, extraímos sinais biológicos do espaço de cores RGB, espaço de cores YUV e dimensão de crominância.

(1) Dimensão do espaço de cores . Dividimos o bloco em três canais de RGB (YUV) e calculamos a média de pixels de cada canal para todos os blocos. Então, 3 sequências de comprimento m n podem ser derivadas em um quadro. Enquanto isso, o mesmo bloco em quadros diferentes também é alterado com quadros. Quando esses procedimentos são empregados em quadros *T* , podemos obter uma matriz tridimensional com o formato *T N* 3, onde *T* denota o número de quadros, *N* denota o número de blocos e 3 representa três canais (RGB ou YUV). Cada linha da matriz representa a alteração do mesmo bloco em diferentes quadros, e cada coluna representa as alterações de diferentes blocos no mesmo quadro.

(2) Dimensão de Crominância . Calculamos a crominância média de cada bloco [5]. Para cada pixel, o sinal de crominância C é calculado como a combinação linear de dois sinais Xf e Yf:onde σ (Xf) e σ (Yf) denotam os desvios padrão de Xf e

Yf. Os sinais Xf e Yf são sinais filtrados passa-banda obtidos, respectivamente, dos sinais X e Y,onde Rn, e Bn são os valores normalizados dos canais de cores individuais. Quando adotamos as operações para todos os blocos e quadros T , podemos obter uma matriz bidimensional com a forma $T\ N$.

Essas matrizes são armazenadas como mapas de cores (matriz tridimensional) e mapas em escala de cinza (matriz bidimensional) para formar o ppg_map correspondente. Em seguida, movemos a janela deslizante para gerar o próximo ppg_map da mesma forma.

3.3. Classificação baseada em CNN

Usamos um classificador CNN para classificar o ppg_map gerado. A rede consiste em seis camadas convolucionais, utilizando a função de ativação 'relu', seguida por uma camada achatada. Existem duas camadas totalmente conectadas após as camadas convolucionais. A última camada totalmente conectada usa 'softmax' como função de ativação e gera as pontuações das classes positivas e negativas. Para evitar overfitting, adicionamos uma camada de dropout, conforme mostrado na Figura 5 .

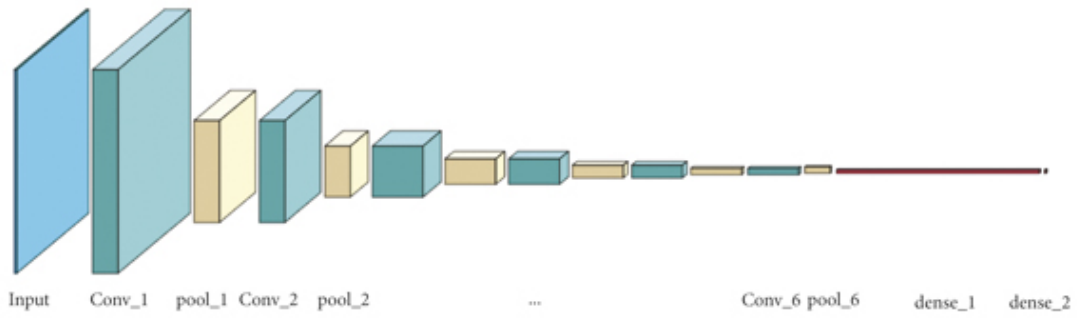


Figura 5

Arquitetura CNN. Usamos seis camadas de convolução com pooling máximo, seguidas por uma camada achatada e camadas densas.

Para cada dimensão na Seção 3.2.3 , treinamos o modelo e obtemos a precisão do conjunto de teste. Além disso, combinamos os sinais das três dimensões para tomar a decisão final.

4. Resultados

Nesta seção, apresentaremos os detalhes de nosso experimento. Primeiro, descrevemos o conjunto de dados que usamos. Em seguida, fornecemos

configurações experimentais detalhadas e o resultado do experimento.

4.1. Conjunto de dados

Usamos três conjuntos de dados públicos para treinar e testar nosso método. Para cada conjunto de dados, geramos o ppg_maps e o dividimos em conjunto de treinamento, conjunto de validação e conjunto de teste de acordo com a proporção de 6: 2: 2. Otimizamos nosso modelo no conjunto de treinamento e no conjunto de validação e, em seguida, obtemos a precisão forense no conjunto de teste.

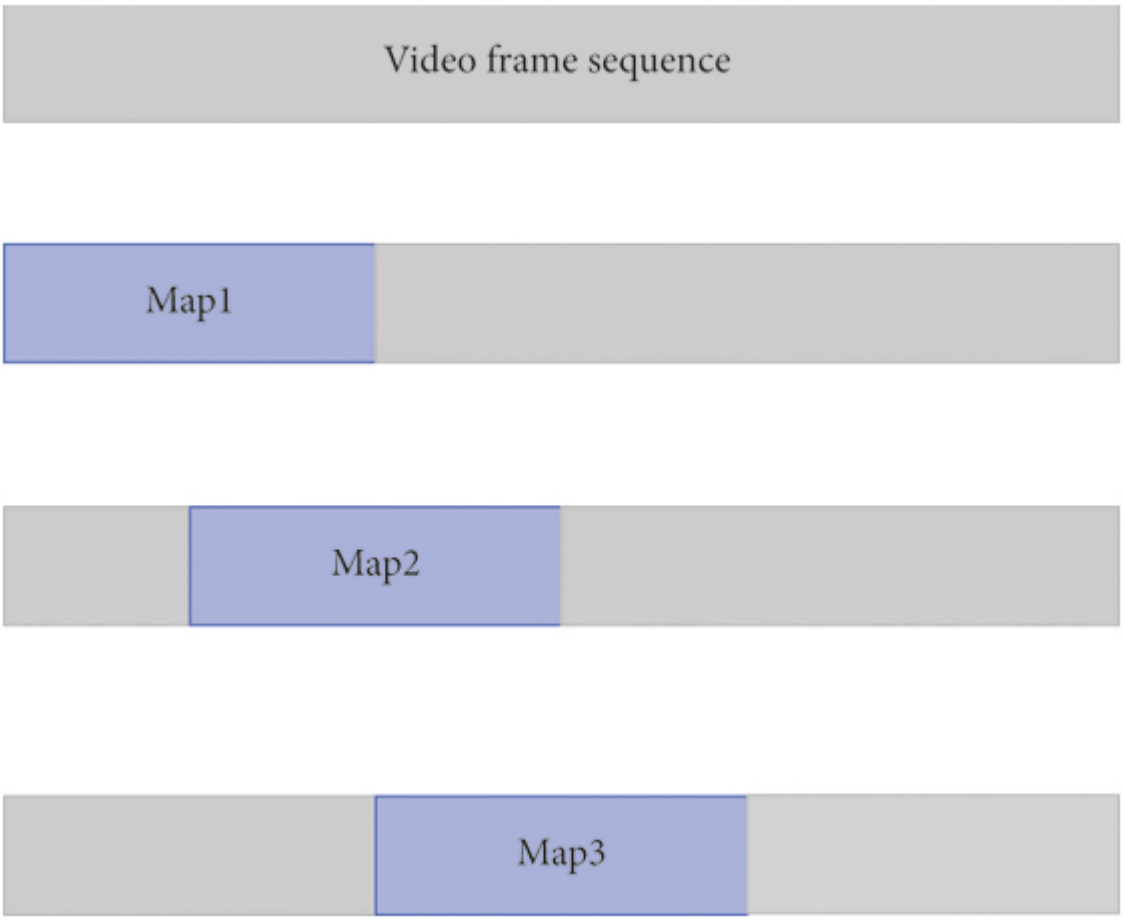
4.1.1. Perícia Facial++

O conjunto de dados FF++ é proposto por Andreas [26], consistindo em 1000 sequências de vídeo originais que foram manipuladas com quatro métodos automatizados de manipulação facial: Deepfakes, Face2Face, FaceSwap e NeuralTextures. Os dados foram provenientes de 977 vídeos do YouTube, e todos os vídeos contêm uma face frontal rastreável, em sua maioria, sem oclusões, o que permite que métodos automatizados de adulteração gerem falsificações realistas. Devido ao método Face2Face e NeuralTextures no conjunto de dados FF++ não adulterar todo o rosto (obtemos sinais biológicos de todo o rosto e quando a parte adulterada é muito pequena, reduzirá a eficácia do método), verificamos principalmente nosso método em conjuntos de dados Deepfakes e FaceSwap.

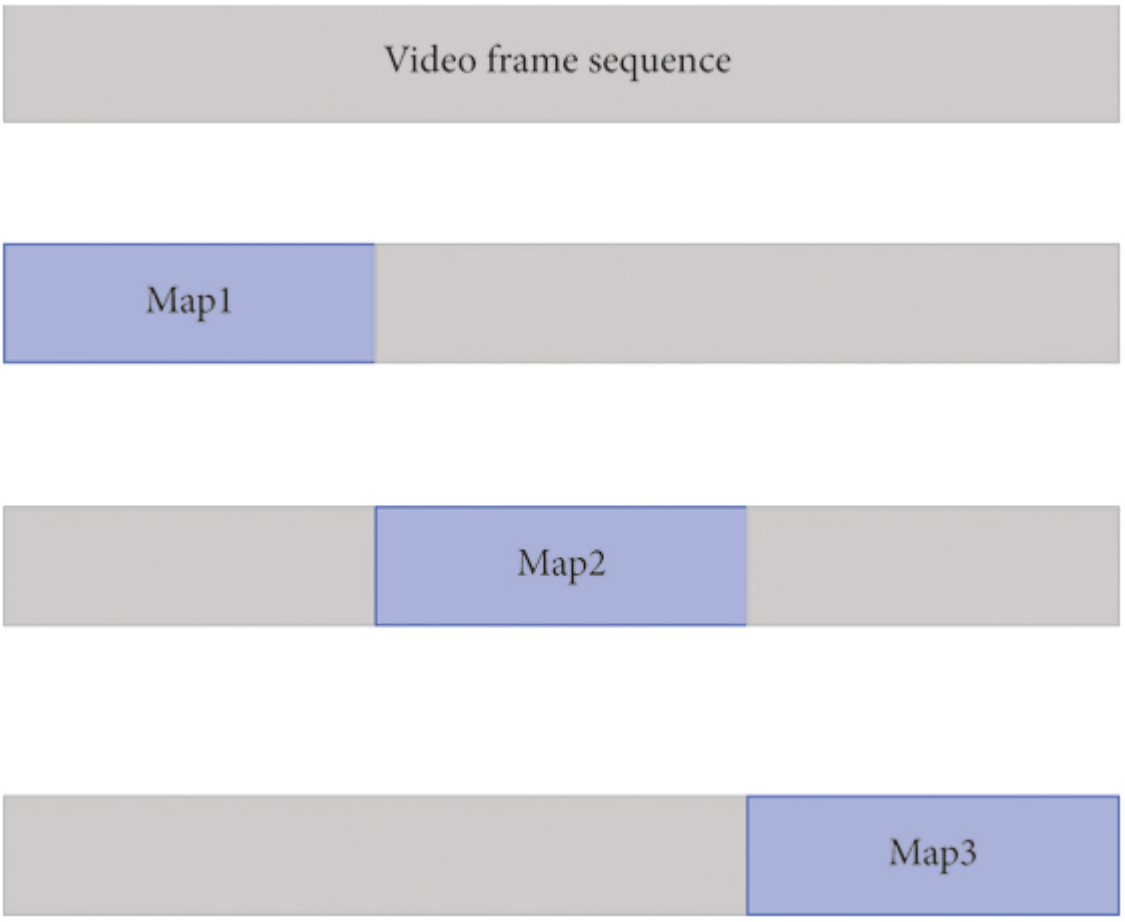
4.1.2. Detecção DeepFake

O conjunto de dados DFD contém 363 vídeos originais realizados por atores e 3.068 vídeos manipulados. Esses atores são obrigados a realizar diferentes ações e, em seguida, implementar a tecnologia de troca facial entre os diferentes atores. Para melhor extrair os sinais biológicos do rosto, escolhemos algumas ações específicas, como “fala feliz no pódio” e “falando quieto”. Nessas ações, o rosto fica bem voltado para a câmera e não há muitos fatores de interferência. Portanto, utilizamos 176 vídeos reais e 754 vídeos falsos do DFD. O maior problema com o conjunto de dados DFD é o desequilíbrio de amostras positivas e negativas. Então, devemos expandir o vídeo real. O princípio da expansão é que qualquer segmento do vídeo real também é um vídeo real. Então usamos a ideia de janela deslizante para gerar o ppg_map. Ao processar vídeo real,6 . Após a expansão, equivalemos a usar 704 vídeos reais e 754 vídeos falsos.

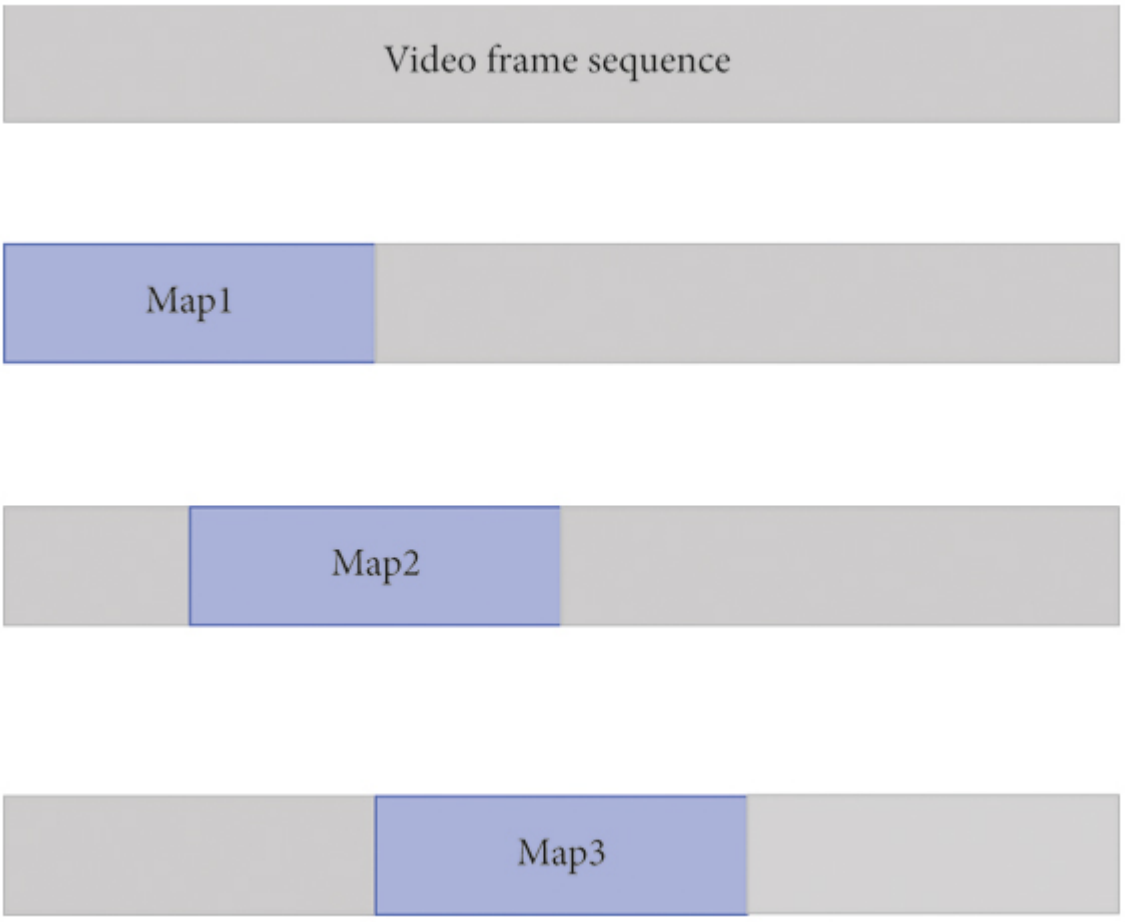
(b)



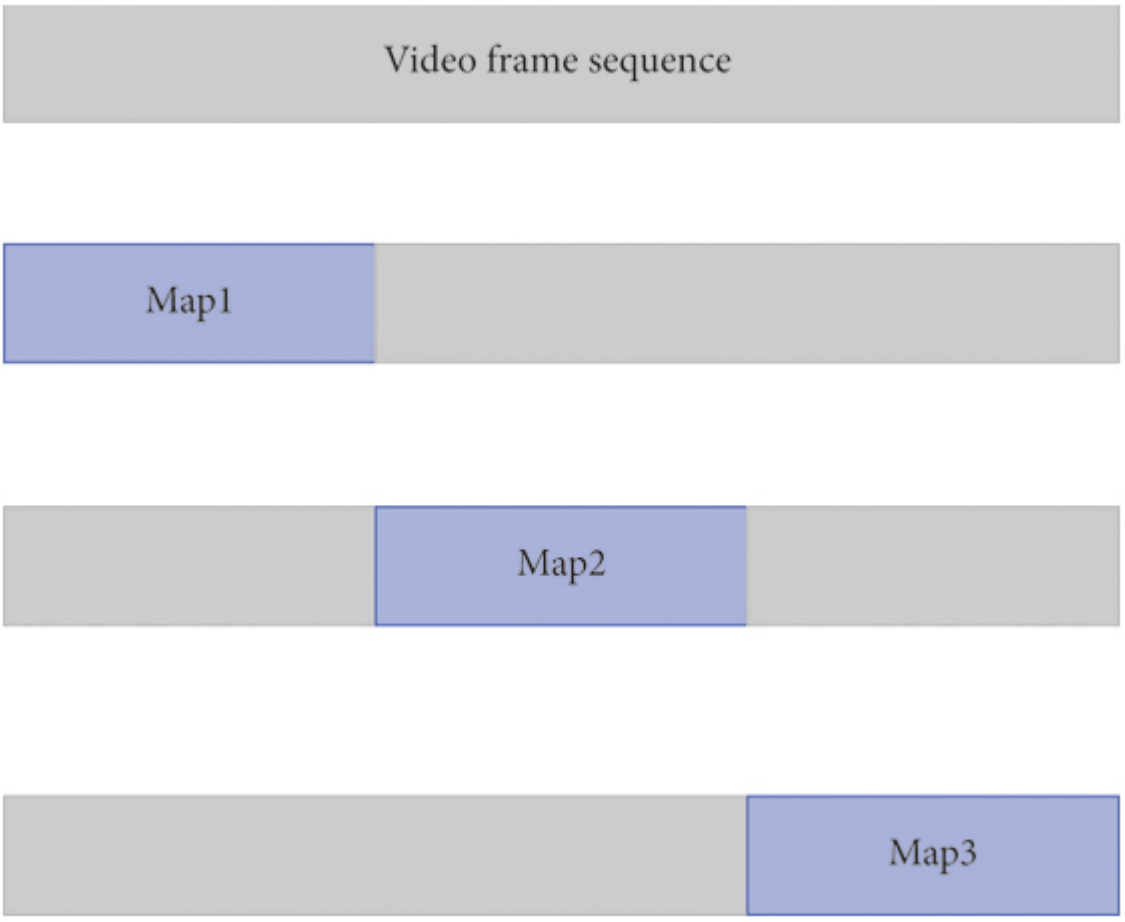
(a)



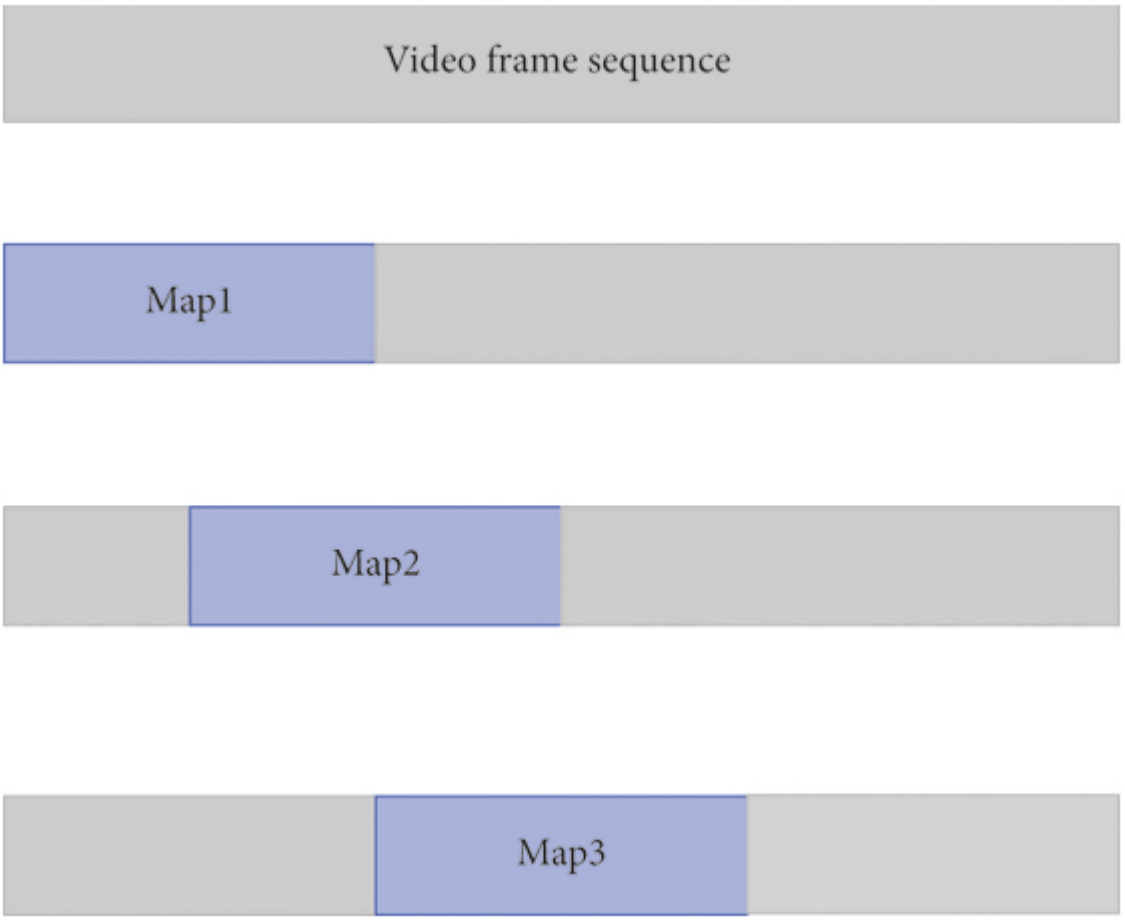
(b)



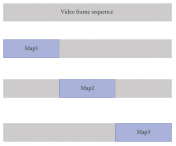
(a)



(b)



- **(a)**



- **(b)**

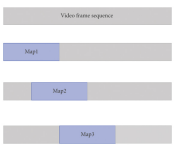


Figura 6

Diferentes maneiras de gerar ppg_map para vídeos reais e falsos. (a) Ao lidar com vídeo falso, a largura da janela deslizante é igual ao comprimento da janela

deslizante. (b) Ao lidar com um vídeo real, o avanço da janela deslizante é menor que o comprimento da janela deslizante.

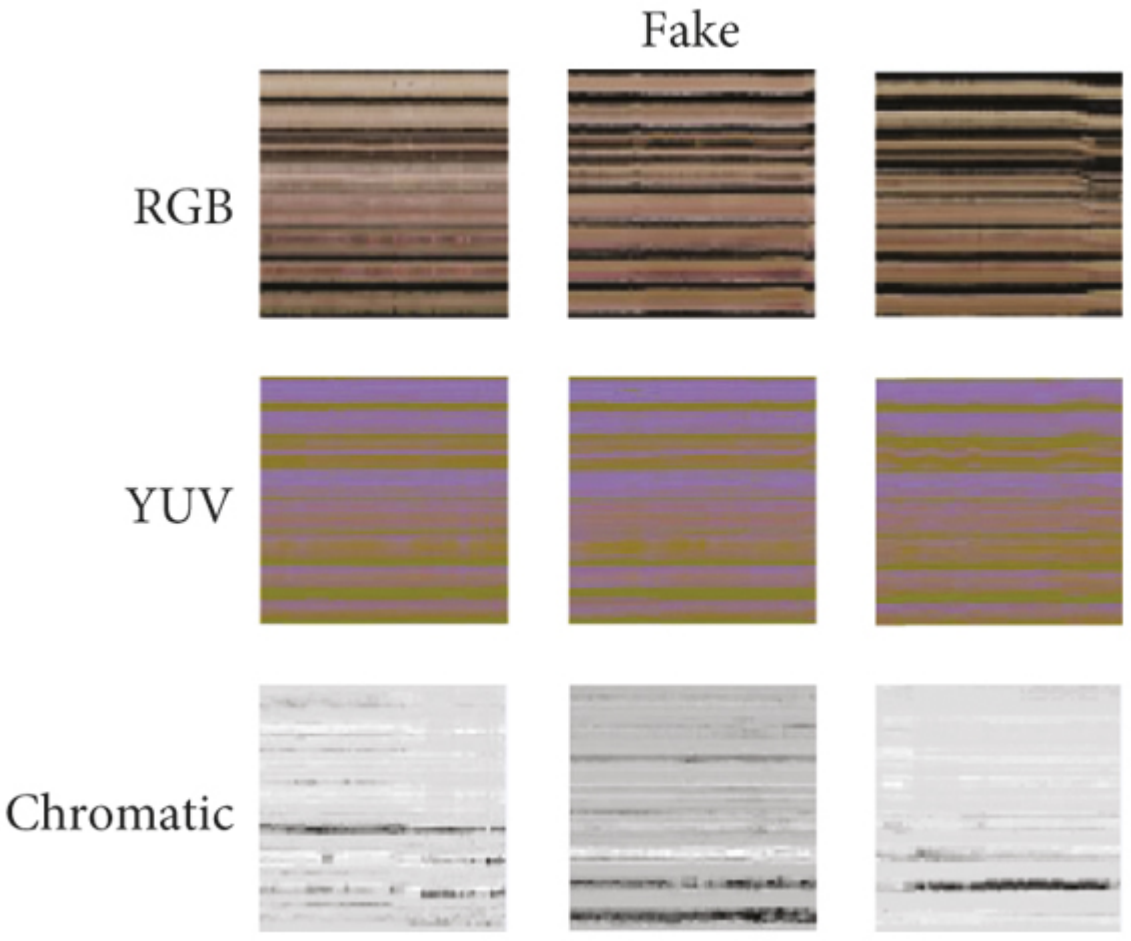
4.1.3. UADFV

O conjunto de dados UADFV é proposto por Yang et al. [21], que contém 49 vídeos reais e 49 vídeos falsos. A duração média de cada vídeo é de cerca de 11 segundos e a resolução é de 294.500 pixels.

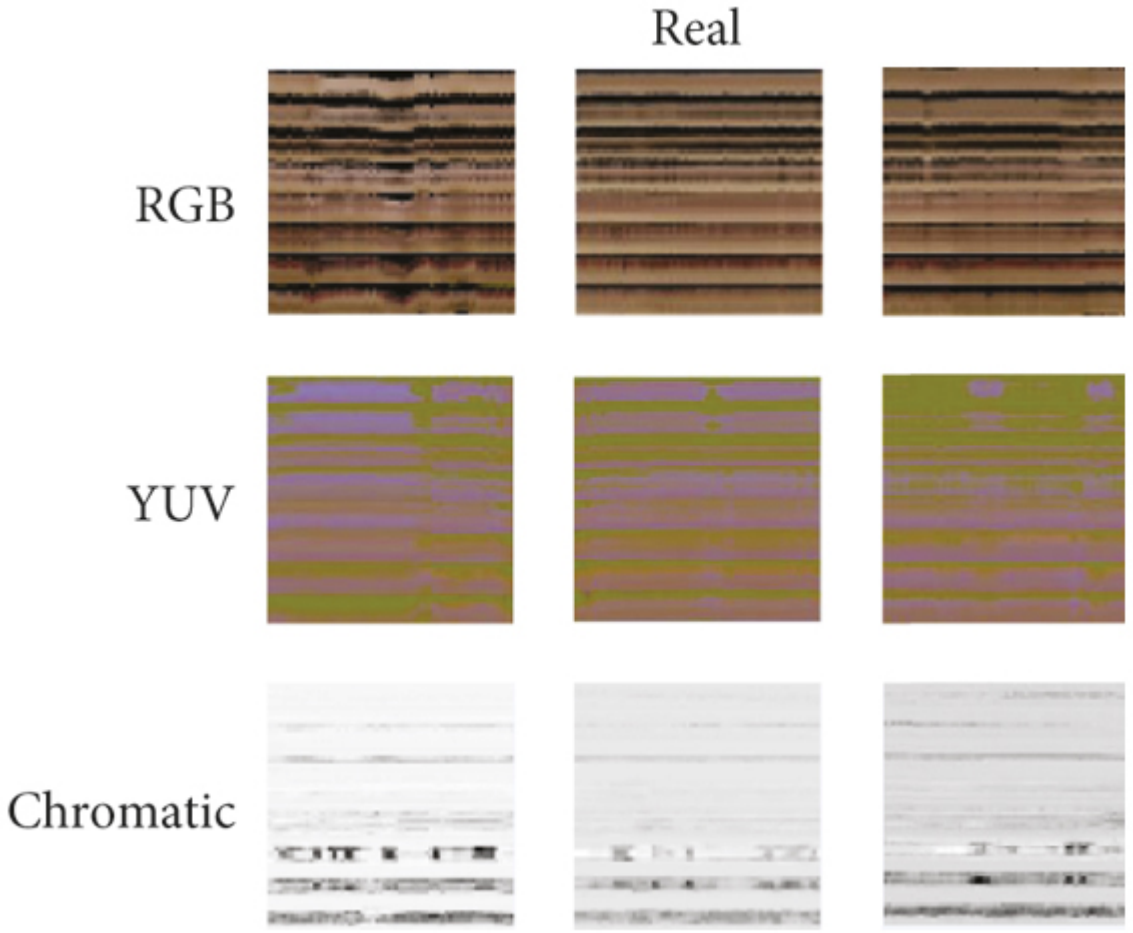
4.2. Configuração e resultados do experimento

Para gerar ppg_map, dividimos o quadro facial em 8 8 blocos (N = 64) e usamos 64 quadros (T = 64) para gerar um ppg_map (o que significa que o comprimento da janela deslizante é 64), então os pixels de cada ppg_maps são 64 64. A Figura 7 mostra um esquema do ppg_map.

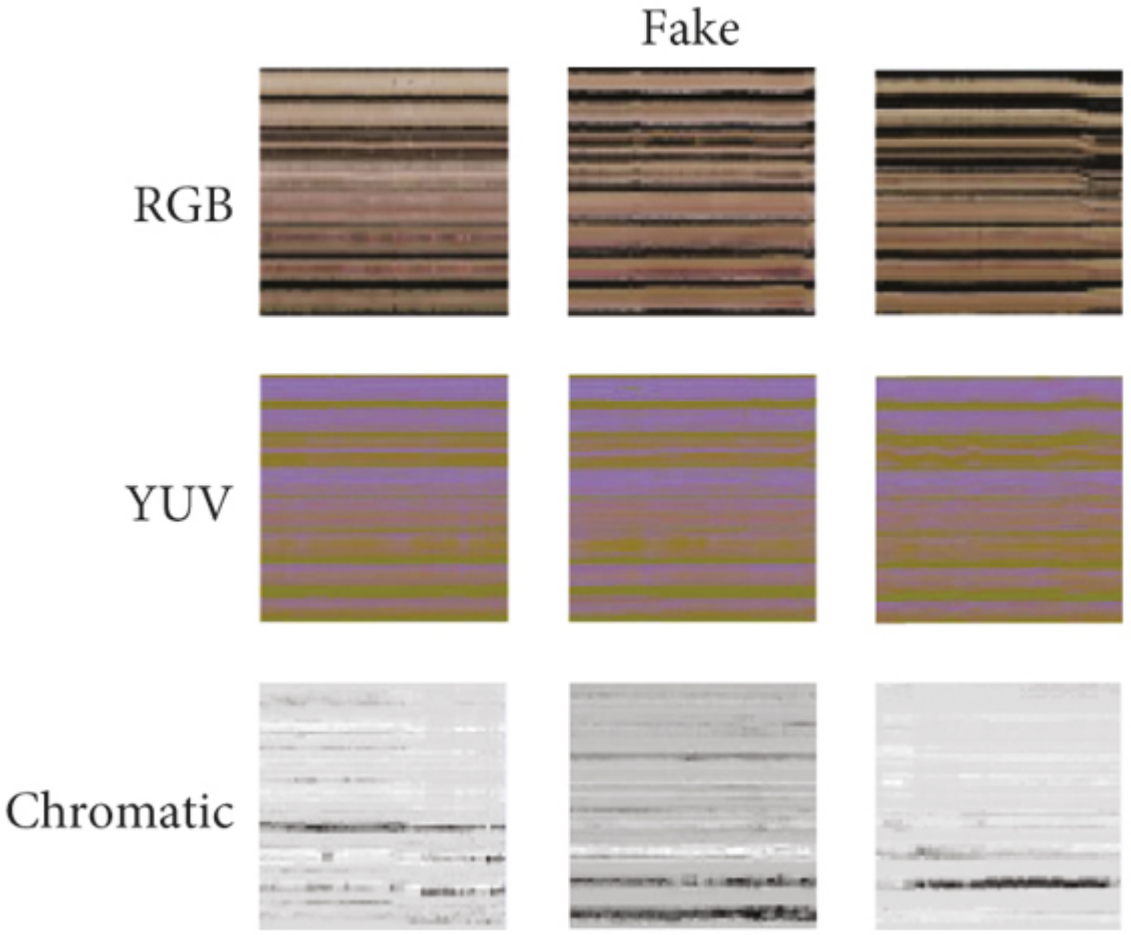
(b)



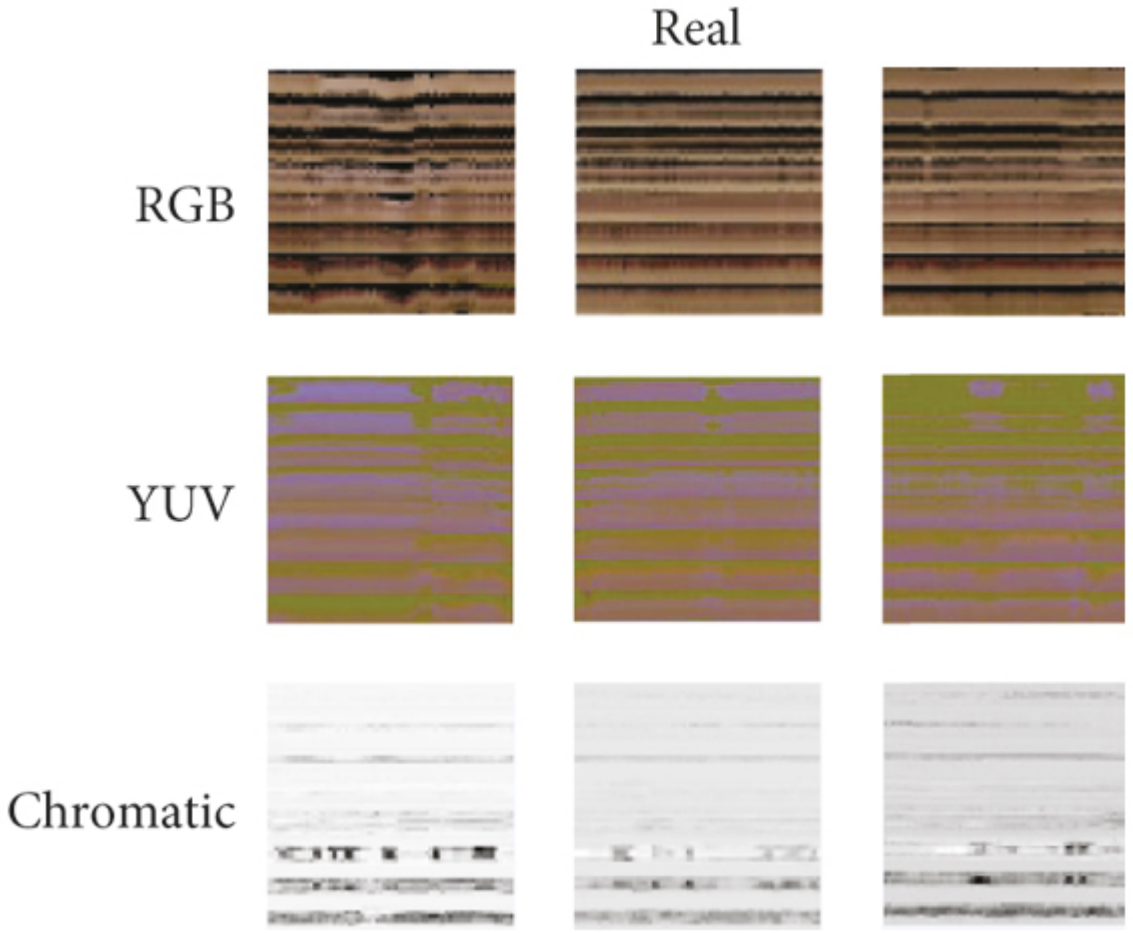
(a)



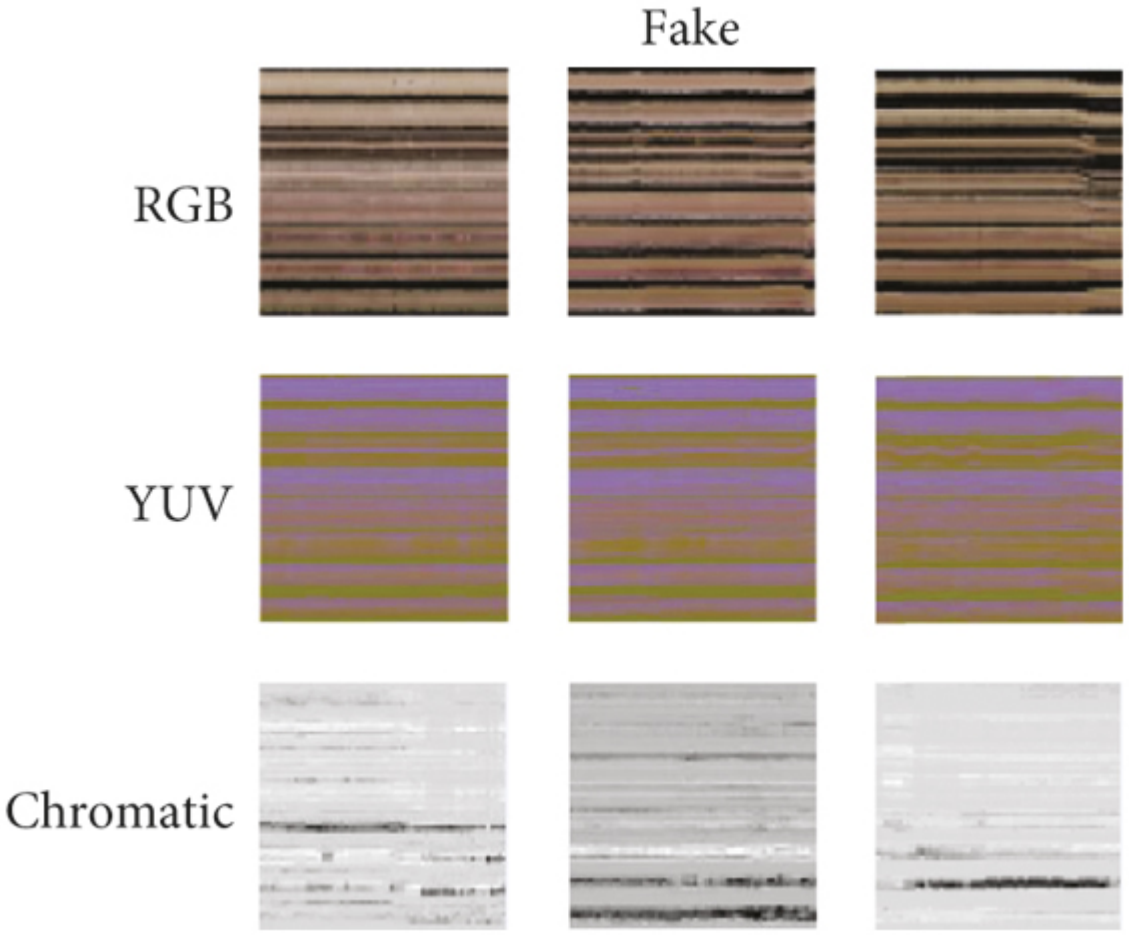
(b)



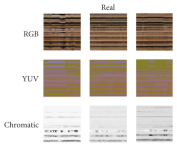
(a)



(b)



- **(a)**



- **(b)**

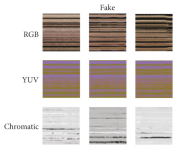


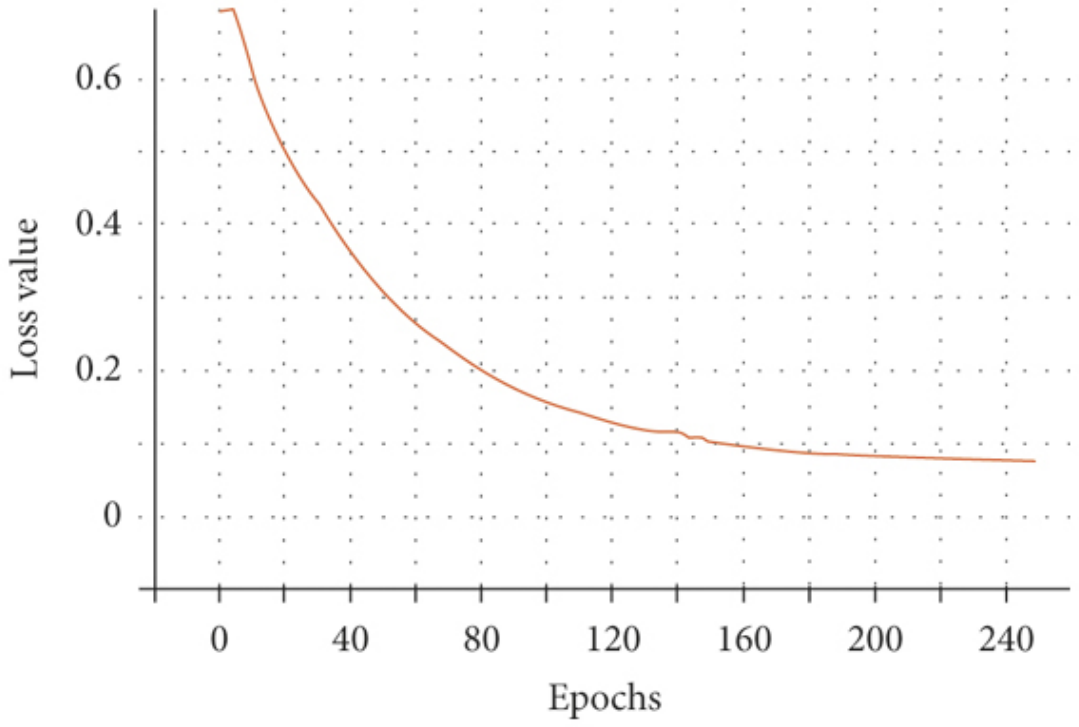
Figura 7

Diagrama esquemático do ppg_map. (a) Os ppg_maps gerados por vídeos reais. (b) Os ppg_maps gerados por vídeos falsos.

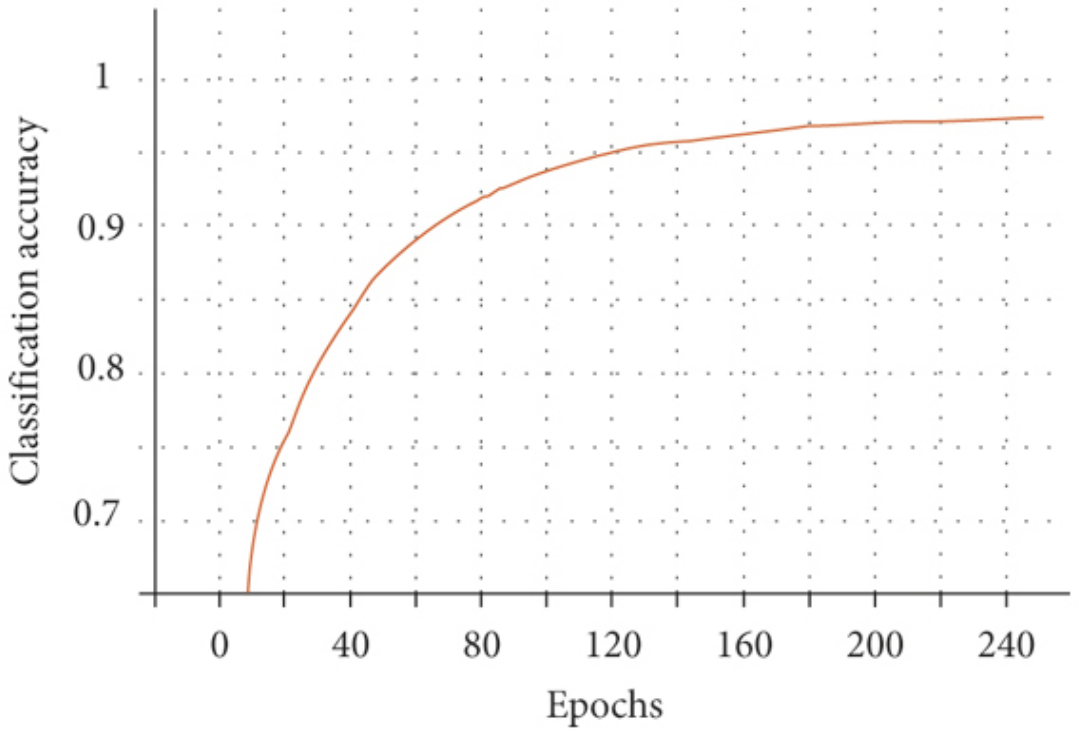
Implementamos esse código em uma estação de trabalho com quatro placas GPU 2080Ti. O modelo foi treinado usando RMSprop por 160 épocas com taxa de aprendizado de 0,0004.

Usamos o conjunto de dados Deepfakes em FF++ (dimensão RGB) para verificar a eficácia do modelo. Os valores de precisão e perda deste modelo no conjunto de treinamento são mostrados na Figura 8 . Pode-se observar na Figura 8 que à medida que as épocas aumentam, a precisão da classificação do modelo aumenta gradativamente, enquanto o valor da perda diminui gradativamente e se estabiliza em 160 épocas, o que ilustra a eficácia do modelo neste artigo.

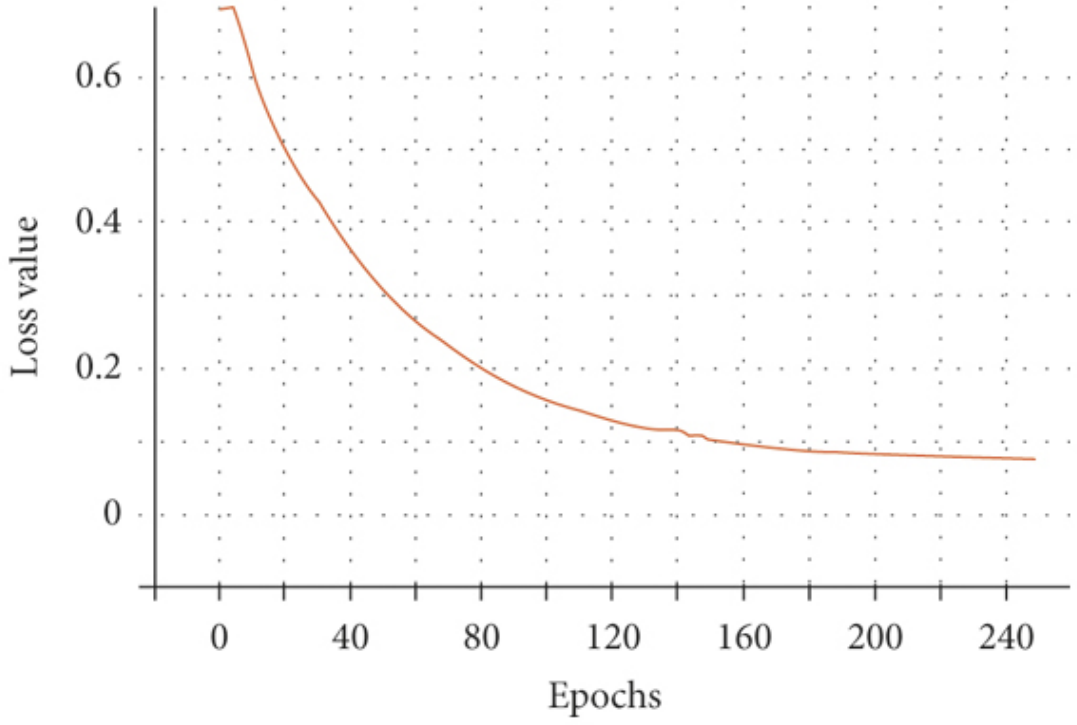
(b)



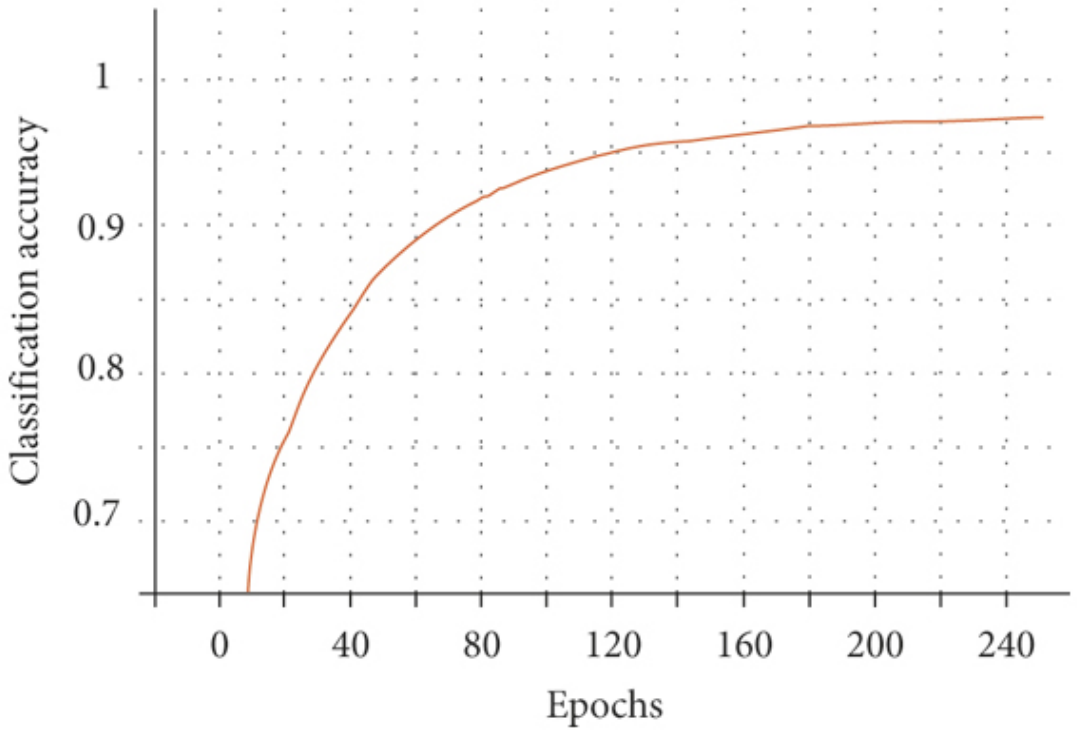
(a)


Classification accuracy
1
0.9
0.8
0.7
0
40
80
120
160
200
240
Epochs

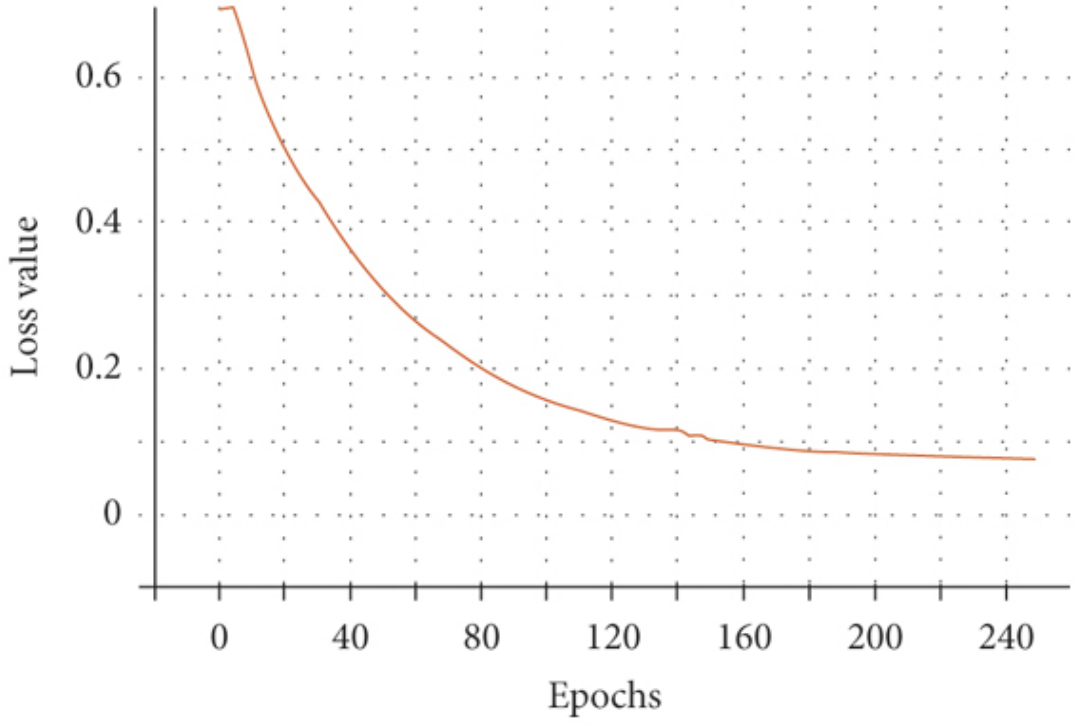
(b)


Loss value
0.6
0.4
0.2
0
0
40
80
120
160
200
240
Epochs

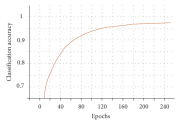
(a)


Classification accuracy
1
0.9
0.8
0.7
0
40
80
120
160
200
240
Epochs

(b)


Loss value
0.6
0.4
0.2
0
0
40
80
120
160
200
240
Epochs

- **(a)**



- **(b)**

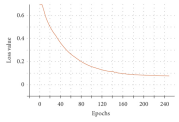


Figura 8

((a), (b)) A curva variável das taxas de precisão e valor de perda com os tempos de treinamento, respectivamente.

Para comprovar a vantagem dos sinais multidimensionais, analisamos a precisão da classificação de sinais unidimensionais e sinais multidimensionais, conforme mostrado na Tabela 1 . A precisão pode ser melhorada obviamente ao usar sinais multidimensionais (MD).

tabela 1

A precisão do conjunto de testes em diferentes conjuntos de dados.

4.3. Comparação

Para verificar a eficácia do método, foi realizado um experimento comparativo com o modelo citado no FaceForensics++, e os resultados da comparação são apresentados na Tabela 2 . Os resultados mostram que nosso método possui maior precisão de detecção do que outros métodos.

mesa 2

Comparação dos resultados de precisão experimental por diferentes modelos.

5. Conclusões

Neste artigo, propomos um método forense baseado em sinais biológicos, através de uma rede neural profunda para realizar a classificação de vídeos reais e falsos. O deepfake não consegue reter com eficácia os sinais biológicos no vídeo facial. Consequentemente, utilizamos sinais biológicos multidimensionais para analisar as diferenças entre vídeos reais e falsos. No entanto, alguns vídeos deepfake são difíceis de serem expostos em condições complicadas, como movimentos instáveis de personagens e mudanças complexas de cena. Esperamos

que a detecção de deepfake nesses cenários possa ser resolvida de forma eficaz usando aprimoramento de sinal e remoção de ruído em um trabalho futuro próximo.

Disponibilidade de dados

Os dados utilizados para apoiar as conclusões deste estudo estão incluídos no artigo.

Conflitos de interesse

Os autores declaram não ter conflitos de interesse.

Agradecimentos

Este trabalho foi parcialmente apoiado pela Fundação Nacional de Ciências Naturais da China NSFC (números de concessão 62072343, U1736211), o Programa Nacional de Desenvolvimento de Pesquisa Chave da China (números de concessão 2019QY (Y) 0206). As opiniões e conclusões aqui contidas são de responsabilidade dos autores e não devem ser interpretadas como representando necessariamente as políticas ou endossos oficiais.

Referências

1. P. Korshunov e S. Marcel, “Deepfakes: uma nova ameaça ao reconhecimento facial? avaliação e detecção”, 2018, https://arxiv.org/abs/812.08685 .Veja em: Google Acadêmico

2. UA Ciftci, I. Demir e L. Yin, “Fakecatcher: detecção de vídeos de retratos sintéticos usando sinais biológicos”, 2020, http://arxiv.org/abs/1901.02212 .Veja em: Google Acadêmico

3. V. Conotter, E. Bodnari, G. Boato e H. Farid, “Detecção com base fisiológica de rostos gerados por computador em vídeo”, em *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)* , pp. , IEEE, Paris, França, outubro de 2014.Veja em: Google Acadêmico

4. PV Rouast, MTP Adam, R. Chiong, D. Cornforth e E. Lux, “Medição remota da frequência cardíaca usando vídeo facial RGB de baixo custo: uma revisão da literatura técnica”, Frontiers of *Computer Science* , vol. 12, não. 5, pp. 858–872, 2018.Ver em: Site do Editor | Google Scholar

5. S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, JF Cohn e N. Sebe, “Completamento de matriz auto-adaptável para estimativa de frequência cardíaca a partir de vídeos faciais sob condições realistas”, em Proceedings of the IEEE *Conference sobre visão computacional e reconhecimento de padrões* , pp. 2396–2404, Las Vegas, NV, EUA, junho de 2016.Veja em: Google Acadêmico

6. G. Balakrishnan, F. Durand e J. Guttag, “Detecting pulse from head motions in

video," em *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* , pp. 3430–3437, Portland, OR, EUA, junho de 2013 .Veja em: Google Acadêmico
7. X. Niu, S. Shan, H. Han e X. Chen, "Rhythmnet: estimativa de frequência cardíaca ponta a ponta do rosto por meio de representação espaço-temporal", *IEEE Transactions on Image Processing* , vol. 29, pp.Veja em: Google Acadêmico
8. B. Chen, W. Tan, G. Coatrieux, Y. Zheng e YQ Shi, "Um esquema de localização de falsificação de cópia e movimento de imagem serial com distinção de origem/destino", *IEEE Transactions on Multimedia* , p. 1, 2020.Ver em: Site do Editor | Google Scholar
9. B. Chen, X. Qi, Y. Zhou, G. Yang, Y. Zheng e B. Xiao, "Localização de emenda de imagem usando imagem residual e rede totalmente convolucional baseada em resíduo", Journal of Visual Communication and Image Representation, *vol* . . 73, Artigo ID 102967, 2020.Ver em: Site do Editor | Google Scholar
10. HH Nguyen, J. Yamagishi e I. Echizen, "Cápsula forense: usando redes de cápsulas para detectar imagens e vídeos forjados", em *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , pp. 2307–2311, IEEE, maio de 2019.Veja em: Google Acadêmico
11. NT Do, IS Na e SH Kim, "Detecção de rosto forense de gans usando rede neural convolucional", 2018.Veja em: Google Acadêmico
12. D. Afchar, V. Nozick, J. Yamagishi e I. Echizen, "Mesonet: uma rede compacta de detecção de falsificação de vídeo facial", em Proceedings *of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS)* , pp. 7, IEEE, Hong Kong, China, dezembro de 2018.Veja em: Google Acadêmico
13. N. Bonettini, ED Cannas, S. Mandelli, L. Bondi, P. Bestagini e S. Tubaro, "Detecção de manipulação de rosto de vídeo por meio de conjunto de CNNs", 2020, http://arxiv.org/abs/2004.07676 .Veja em: Google Acadêmico
14. Y. Li e S. Lyu, "Expondo vídeos deepfake detectando artefatos de distorção facial", 2018, http://arxXiv.org/abs/:1811.00656 .Veja em: Google Acadêmico
15. H. Zhao et al., "Detecção de deepfake multiatencional", pré-impressão arXiv arXiv:2103.02406 (2021).Veja em: Google Acadêmico
16. H. Liu, W. Zhou, D. Chen, T. Wei, W. Zhang e N. Yu, "Aprendizagem superficial em fase espacial: repensando a detecção de falsificação facial no domínio da frequência", 2021, http://arxXiv.org /abs/:2103.01856 .Veja em: Google Acadêmico
17. D. Güera e EJ Delp, "Deepfake video Detection using Recurrent Neural Networks", em *Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* , pp. 2018, novembro.Veja

em: Google Acadêmico
18. E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi e P. Natarajan, “Estratégias convolucionais recorrentes para detecção de manipulação facial em vídeos”, *Interfaces (GUI)* , vol. 3, não. 1, 2019.Veja em: Google Acadêmico
19. S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano e H. Li, “Protegendo líderes mundiais contra falsificações profundas”, em Proceedingd *of the CVPR Workshops* , pp. , 2019, junho.Veja em: Google Acadêmico
20. Y. Li, MC Chang e S. Lyu, “Ictu oculi: expor ai criou vídeos falsos detectando piscar de olhos”, em *Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS)* , pp. IEEE, Hong Kong, China, 2018, dezembro.Veja em: Google Acadêmico
21. X. Yang, Y. Li e S. Lyu, “Expondo falsificações profundas usando poses de cabeça inconsistentes”, em *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , pp. , IEEE, Brighton, Reino Unido, 2019, maio.Veja em: Google Acadêmico
22. L. Li, J. Bao, T. Zhang et al., “Raio-x facial para detecção mais geral de falsificação de rosto”, em *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* , pp. , WA, EUA, agosto de 2020.Veja em: Google Acadêmico
23. R. Wang, F. Juefei-Xu, L. Ma e X. Xie, “FakeSpotter: uma linha de base simples, mas robusta para detectar rostos falsos sintetizados por IA”, 2019, http://arxXiv.org/abs/:1909.06122 .Veja em: Google Acadêmico
24. SKA Prakash e CS Tucker, “Método de filtro Bounded Kalman para estimativa de frequência cardíaca sem contato e robusta ao movimento”, *Biomedical Optics Express* , vol. 9, não. 2, pp. 873–897, 2018.Ver em: Site do Editor | Google Scholar
25. P. Viola e M. Jones, “Detecção rápida de objetos usando uma cascata reforçada de recursos simples”, em *Anais da conferência da sociedade de computação IEEE de 2001 sobre visão computacional e reconhecimento de padrões. CVPR* , IEEE, Kauai, HA, EUA, 2001, dezembro.Veja em: Google Acadêmico
26. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies e M. Nießner, “Faceforensics: um conjunto de dados de vídeo em grande escala para detecção de falsificação em rostos humanos”, 2018, http://arxXiv . org/abs/:1803.09179 .Veja em: Google Acadêmico
27. B. Bayar e MC Stamm, “Uma abordagem de aprendizagem profunda para detecção universal de manipulação de imagens usando uma nova camada convolucional”, em *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security* , pp. , Junho.Veja em: Google Acadêmico
28. N. Rahmouni, V. Nozick, J. Yamagishi e I. Echizen, “Distinguindo computação

gráfica de imagens naturais usando redes neurais de convolução”, em Proceedings *of the IEEE Workshop on Information Forensics and Security (WIFS)* , IEEE, Rennes, França , janeiro de 2017.Veja em: Google Acadêmico
29. J.-Y. Baek, Y.-S. Yoo e S.-H. Bae, “Aprendizagem generativa de conjunto adversário para análise forense facial”, *IEEE Access* , vol. 45421–45431, 2020.Ver em: Site do Editor | Google Scholar
30. A. Rossler, D. Cozzollino, L. Verdoliva, C. Riess, J. Thies e M. Niessnar, “Faceforensics++: aprendendo a detectar imagens faciais manipuladas”, em *Anais da Conferência Internacional IEEE/CVF sobre Visão Computacional* , Seul , Coreia, fevereiro de 2019.Veja em: Google Acadêmico
31. N. Dogonadze, O. Jana e Ji Hou, “Detecção de falsificação de rosto profundo”, 2020, http://arxiv.org/abs/2004.11804 .Veja em: Google Acadêmico

direito autoral