VISÃO DA UNDB

Ser uma instituição nacionalmente reconhecida pela sua excelência em todas as suas áreas de atuação.

MISSÃO DA UNDB

Promover o conhecimento alicerçado em princípios éticos, científicos e tecnológicos, através de metodologias de vanguarda, visando à formação e ao aperfeiçoamento humano de profissionais comprometidos com o processo de desenvolvimento e mudança nos seus campos de atuação.

1 INFORMAÇÕES SOBRE O COMPONENTE CURRICULAR

Disciplina: UNDB 4.0 Visão Computacional/PBL	**Carga Horária: 80 h**
Professor: Felipe Gomes	**Turno: Noturno**
Curso: Escola de Tecnologia	**Período/Semestre: 2º/2025.2**

PROBLEMA 2 – DETECÇÃO DE DEEPFAKES EM VÍDEOS INFORMATIVOS

Vídeos ocupam lugar privilegiado na construção do “real”: um rosto que fala e se move diante da câmera costuma ser percebido como prova direta de um fato. É nesse terreno que os deepfakes, conteúdos audiovisuais sintéticos ou manipulados com alto realismo, ganham força, imitando rostos, vozes e expressões com alto poder de persuasão. A dinâmica das plataformas amplia o problema: estudos mostram que notícias falsas tendem a se espalhar mais rápido e mais longe do que as verdadeiras, encurtando a janela para verificação antes de afetar percepções e decisões públicas (VOSOUGHI; ROY; ARAL, 2018). Relatórios independentes também apontam o crescimento acelerado e a democratização das ferramentas de síntese, com o número de vídeos deepfake monitorados dobrando em poucos meses e migrando para serviços acessíveis ao público leigo (DEEPTRACE, 2019).

Do ponto de vista técnico, a popularização de redes adversárias generativas (GANs) tornou viável a síntese realista de faces e movimentos (GOODFELLOW et al., 2014). Em paralelo, bases de referência e estudos sistemáticos documentam como manipulações faciais podem soar plenamente plausíveis para observadores e, até confundir sistemas automáticos, tensionando a confiança histórica no vídeo como “registro factual” (RÖSSLER et al., 2019). Os

impactos extrapolam o entretenimento: análises jurídicas discutem riscos à privacidade, à democracia e à segurança nacional, além de alertar para erros de rotulagem que podem injustamente taxar conteúdos autênticos como falsos (CHESNEY; CITRON, 2019).

Nesse cenário, discutir deepfakes significa refletir sobre credibilidade, ritmo de circulação e responsabilidade na comunicação pública, bem como sobre a necessidade de evidências transparentes e critérios claros de verificação.

Figura 1 - 'Deepfakes' desencadeiam uma corrida para combater fotos e vídeos manipulados

Identificar manipulações em vídeos não exige, necessariamente, técnicas complexas de inteligência artificial. É possível obter bons indícios apenas com métodos determinísticos e explicáveis, que são recursos tradicionais, mas ainda muito eficazes quando aplicados com rigor. O processo começa com algo simples: extrair quadros do vídeo em intervalos regulares (por exemplo, 5 ou 10 por segundo) e preparar essas imagens para análise. Aqui entram etapas como a conversão para espaços de cor mais adequados à detecção de padrões, como HSV, e o ajuste de brilho e contraste. Para isolar a região do rosto (ponto central em vídeos manipulados) pode-se usar segmentação por cor de pele. Quando essa abordagem falha devido a iluminação ou tom de pele, uma alternativa é selecionar a região manualmente no primeiro quadro e rastreá-la ao longo do vídeo usando *template matching*, técnica clássica que compara padrões de pixels entre imagens (BRADSKI; KAHLER, 2008).

Uma vez definida a região de interesse (ROI), é hora de aplicar ferramentas que revelam o que o olho humano talvez não perceba. Detectores de borda como Canny ou Sobel mapeiam as arestas da imagem e ajudam a identificar descontinuidades estranhas, como cortes abruptos na linha do cabelo ou no contorno do queixo. No campo da frequência, a Transformada Rápida de Fourier (FFT) e a Transformada Discreta do Cosseno (DCT) mostram o "esqueleto" espectral da imagem: suavizações exageradas, artefatos de blocagem (*blockiness*) ou padrões de recompressão diferentes entre o rosto e o fundo podem ser sinais de edição (GONZALEZ; WOODS, 2018). Também vale medir estatísticas simples de cor, como médias e variações nos canais RGB ou HSV, e comparar a consistência entre pele e áreas próximas, como pescoço e orelhas. Um truque interessante é aplicar um filtro mediano e observar o residual para destacar *halos* e transições artificiais.

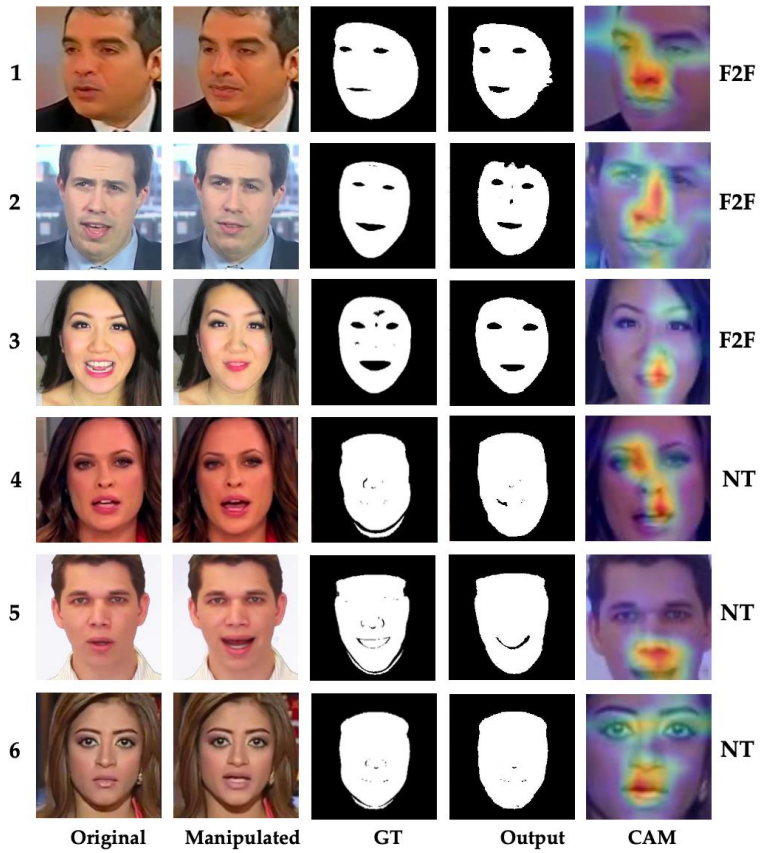

Figura 2 - As colunas 1 e 2 apresentam, respectivamente, as imagens originais e manipuladas. A coluna 3 mostra as máscaras binárias de referência (GT), enquanto a coluna 4 exibe as máscaras previstas e a coluna 5, os mapas de ativação de classe (CAMs) gerados para amostras manipuladas dos conjuntos Face2Face (linhas 1–3) e Neural-Textures (linhas 4–6) (MAZAHERI; ROY-CHOWDHURY, 2022).

Pode-se obter informações bastante relevantes da dimensão temporal. Comparar quadros consecutivos e calcular o fluxo óptico (por exemplo, com o método de Farnebäck (OPENCV, 2025)) permite detectar *flicker* de cor e deslocamentos incoerentes de texturas.

Mesmo sem modelos treinados, é possível acompanhar pequenas áreas do rosto, como olhos e cantos da boca, e registrar mudanças na intensidade ou no contraste. Essa observação pode indicar, por exemplo, piscadas irregulares ou microexpressões “travadas”, comuns em vídeos falsificados.

Por fim, todas essas medições são consolidadas em estatísticas como médias, desvios e percentuais, que servem de base para uma lógica simples: um vídeo é sinalizado como suspeito quando múltiplos indicadores ultrapassam determinados limites. O resultado é um relatório claro e interpretável, onde cada sinal de possível manipulação está documentado.

Para praticar, há acervos públicos valiosos, como FaceForensics++ (https://github.com/ondyari/FaceForensics), o Deepfake Detection Challenge (DFDC) no Kaggle (https://www.kaggle.com/c/deepfake-detection-challenge), Celeb-DF v2 (https://github.com/yuezunli/celeb-deepfakeforensics) e DeeperForensics-1.0 (https://github.com/EndlessSora/DeeperForensics-1.0). Esses conjuntos oferecem vídeos autênticos e manipulados em diferentes condições, perfeitos para colocar essas técnicas “de raiz” à prova.

Questão Norteadora: Diante do aumento da circulação de vídeos manipulados que imitam rostos, vozes e expressões de figuras públicas, de que forma técnicas determinísticas e explicáveis de processamento de imagens, como detecção de bordas, análise espectral, métricas de cor e inspeção temporal, podem ser combinadas para identificar indícios visíveis de adulteração, garantindo transparência no processo de verificação e facilitando a comunicação dos resultados ao público não especializado?

Bibliografias sugeridas:

CHESNEY, R.; CITRON, D. Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*, 2019. Disponível em: https://scholarship.law.bu.edu/faculty_scholarship/640/. Acesso em: 11 ago. 2025.
DEEPTRACE. *The State of Deepfakes: Landscape Report*. 2019. Disponível em: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf. Acesso em: 11 ago. 2025.
GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2014. Disponível em: https://arxiv.org/abs/1406.2661. Acesso em: 11 ago. 2025.

OPENCV. Optical Flow. Disponível em: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html. Acesso em: 16 ago. 2025.

RÖSSLER, A. et al. FaceForensics++: learning to detect manipulated facial images. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019. Disponível em: https://arxiv.org/abs/1901.08971. Acesso em: 11 ago. 2025.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. *Science*, v. 359, n. 6380, p. 1146–1151, 2018. Disponível em: https://www.science.org/doi/10.1126/science.aap9559. Acesso em: 11 ago. 2025.