Detecção de liveness facial de quadro único:

Como funciona para reduzir o atrito e o abandono do usuário

IDR&D
a Mitek company

A capacidade de determinar o liveness facial a partir de uma única imagem permite detecção de ataques de apresentação de alto desempenho sem adicionar atrito à experiência do usuário. Uma abordagem "sem interação" de liveness que não tenha atrito para o usuário elimina o abandono do cliente causado pelas complicações de uma abordagem "ativa". O resultado é menos clientes perdidos e mais receita. Este artigo discute como funciona a detecção de liveness facial com uma única imagem e apresenta um estudo de caso de seu impacto nas operações de onboarding de um banco de varejo.

O que é liveness facial sem interação?

Em processos remotos de verificação de identidade e autenticação que utilizam biometria, a detecção de liveness é fundamental para prevenir *ataques de apresentação*[1], ou "falsificações". Os ataques comuns incluem apresentação de rostos impressos em papel ou exibidos em telas, reproduções de vídeo e máscaras faciais.

As abordagens para detecção de liveness podem ser categorizadas como *ativo* ou *sem interação*. Uma abordagem ativa depende da interação com o usuário, como instruí-lo a piscar, sorrir, virar a cabeça ou mover o dispositivo enquanto estiver diante da câmera e, em seguida, detectar sua reação. O processo leva mais tempo e fornece informações aos fraudadores que podem ser potencialmente usadas para burlar o mecanismo de segurança.

A abordagem de vivacidade passiva ideal usa a mesma selfie para vivacidade que já foi capturada para correspondência de rosto, eliminando os custos de processamento de vídeo ou vários quadros e removendo o atrito adicional do usuário.

Por exemplo, software de animação gratuito facilita a criação de uma curta reprodução de vídeo sorrindo, piscando e virando a cabeça[2]. Usar uma máscara de papel com recortes nos olhos e na boca pode enganar os sistemas de vivacidade ativos. Além disso, uma abordagem ativa requer múltiplos quadros de imagem ou vídeo, exigindo mais processamento e transmissão de dados, acrescentando ainda mais tempo e custo ao processo de autenticação de identidade.

Em contraste, uma técnica de liveness sem interação não requer instruções, comandos ou resposta do usuário. Nem todo liveness sem interação são iguais. Algumas abordagens não dependem da interação do usuário, mas utilizam vídeo ou múltiplas imagens durante o processo de captura, o que tende a exigir mais processamento e transmissão de dados e, portanto, adiciona atrito na forma de latência e custos de infraestrutura computacional e rede.

A abordagem de liveness sem interação ideal usa a mesma selfie que já foi capturada para o matching facial, eliminando os custos de processamento de vídeo ou vários quadros e removendo o atrito adicional do usuário.

ID R&D fornece essa capacidade ideal de liveness sem interação com apenas uma selfiepor meio de seu produto IDLive® Face.

1 Um ataque de apresentação é um método de tentativa de fraude pelo qual um malfeitor apresenta uma falsificação em vez de uma amostra viva e presente. Os exemplos incluem uma cópia impressa, uma imagem em um display digital ou uma máscara.

2 Exemplos incluem MotionPortrait, Deep Nostalgia e Avatarify.

Um resumo do vantagens do liveness com uma única imagem

A seguir está um resumo do vantagens do liveness com apenas uma imagem para usuários e integradores de sistemas de autenticação facial. O resultado é uma solução mais fácil, rápida, segura e menos dispendiosa, que não acrescenta atrito ao usuário e, portanto, não contribui para o abandono do usuário ou para uma experiência negativa.

1. Não há necessidade de educar ou instruir o usuário, nem de movimentos, gestos ou vídeo para reconhecê-lo como uma pessoa viva.

2. A mesma imagem usada para o maching facial é usada para verificação de liveness, reduzindo substancialmente a complexidade do design, suporte e manutenção da solução.

3. Não há necessidade de alterações na interface do usuário ou interfaces de comunicação. Um aplicativo de back-end simplesmente executa uma chamada de API para o IDLive Face implantado na infraestrutura de back-end. Isso torna a integração rápida e direta para os desenvolvedores.

4. Detalhes mínimos são enviados do dispositivo, reduzindo a latência. Uma única imagem geralmente não é maior que 300 kB ao usar as configurações recomendadas da ID R&D em comparação com soluções que enviam vários quadros ou vídeos. Se você já estiver enviando a imagem para o servidor, não haverá sobrecarga adicional para transmissão de dados.

5. A experiência do usuário não fornece informações aos fraudadores sobre como burlar o mecanismo de segurança. Não há uma etapa de ativação separada do matching facial que o fraudador possa atacar.

Ao usar a detecção de liveness sem interação de apenas uma imagem, o usuário não sabe que uma verificação de liveness está acontecendo, oferecendo uma experiência simples para os usuários e simplificada para os desenvolvedores:

- Nenhuma instrução, comando ou resposta para o usuário seguir
- A mesma selfie usada para matching e liveness
- Suporta navegador ou aplicativos nativos
- Integração, manutenção e suporte de software altamente simplificados
- Transferência mínima de dados dispositivo-servidor
- Nenhuma informação para ajudar os fraudadores

Como funciona o liveness com apenas uma imagem

Como é possível avaliar com tanta precisão o liveness a partir de uma única imagem, quando outras soluções utilizam fluxos de vídeo completos que capturam o movimento e a interação do usuário?

A resposta curta é que liveness de imagem única usa IA. Mas cavando mais fundo, a primeira coisa a entender sobre liveness de apenas uma imagem é que existem características de uma imagem digital que – embora não seja facilmente visível a olho nu – pode ser detectado e medido usando visão computacional. Uma analogia é a diferença entre avaliar a qualidade da imagem da sua TV em casa versus em uma loja de eletrônicos, onde muitos estão ligados e mostram o mesmo conteúdo e assim as diferenças entre TVs são fáceis de detectar. A visão computacional pode detectar e medir com precisão essas diferenças usando matemática.

Mas como funciona a IA (versão resumida)? Inteligência artificial é um termo genérico que abrange várias técnicas de aprendizado de máquina, muitas das quais envolvem o treinamento de um *rede neural* profunda para atuar como classificador[3]. Essas redes neurais podem ser pensadas como uma matriz de equações organizadas de modo que as saídas de cada coluna de equações forneçam entradas para múltiplas colunas de equações em série (o que as torna uma rede "profunda").

O processo de treinamento envolve o uso fundamental de dados reais de entrada e saída – essencialmente dados de entrada que são marcados manualmente com dados de saída de "resposta certa" – para treinar a rede qual é a saída correta para uma determinada entrada. As entradas e saídas associadas são utilizadas no processo de treinamento para gerar um conjunto de coeficientes para as equações da rede neural. Diferentes coeficientes são tentados até que a taxa de erro geral é minimizada, o que poderia exigir milhões de cálculos e várias iterações. É um processo de tentativa e erro de força bruta no qual os computadores são muito bons. Uma vez treinada, a rede neural utiliza um conjunto de coeficientes que minimiza o erro e agora pode classificar dados de entrada nunca vistos antes.

3 Um classificador é um tipo de software projetado para categorizar automaticamente os dados de entrada. Um dos primeiros classificadores baseados em aprendizado de máquina foi usado pelos Correios para ler os dígitos de códigos postais manuscritos.

Redes Neurais na Prática

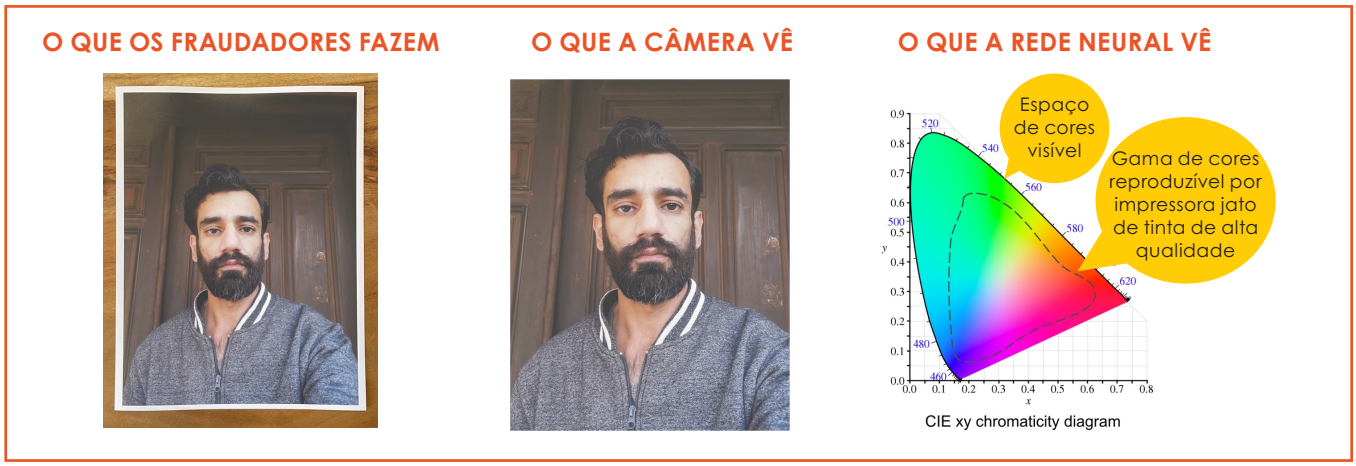
Detectando ataques de apresentação de “repetição de tela”

Os ataques de apresentação de repetição de tela ocorrem quando uma foto exibida em uma tela digital é apresentada como prova de identidade em vez de uma selfie ao vivo. Embora seja difícil de ver a olho nu, há evidências do ataque reveladas nesta única imagem usando técnicas de visão computacional.

O “Efeito Porta de Tela” é o que acontece quando uma foto de uma tela digital é tirada em alta resolução. Embora muito pequena, existe uma grade de espaços entre os pixels da tela que pode ser visto sob ampliação e detectado por uma câmera de alta resolução

Detectando ataques de apresentação de “cópia impressa”

A gama de cores de um dispositivo de impressão é determinada pelo matiz, saturação e luminosidade de suas tintas (ciano, magenta, amarelo, preto), bem como pelo brilho e outras características do substrato em que estão impressos. As cores visíveis ao olho humano têm uma gama de cores maior do que dispositivos de impressão que usam tintas CMYK, especialmente em azuis e pretos profundos. Imprimir uma imagem capturada por uma câmera digital requer a transformação da imagem do espaço de cores RGB da câmera para o da impressora que é o CMYK. Durante este processo, as cores do RGB que estão fora da gama devem ser convertidas para valores aproximados dentro da gama de espaço CMYK. Por estas razões, no impresso a imagem será menos vívida do que a imagem original capturada pela câmera. Quando uma cor está “fora de gama”, ele não pode ser convertido fielmente para o dispositivo de destino.

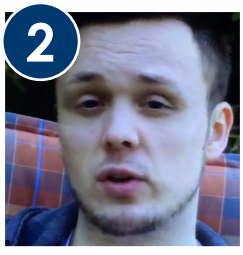


Redes Neurais – uma analogia física

Uma analogia física de uma rede neural é Plinko[4], um jogo jogado no game show "The Price Is Right" por quarenta anos, onde um disco redondo cai em uma placa vertical cheia de alfinetes; a bola atinge os pinos como ele cai, viajando aleatoriamente e então pousando em uma das várias caixas numeradas na parte inferior que têm prêmios diferentes para o jogador.

Em um jogo Plinko normal, a rota e o fim do disco é aleatório. Mas e se, por tentativa e erro, poderíamos organizar com precisão os pinos de modo que o disco sempre pousaria em uma caixa específica com base nas suas características físicas; digamos, o tamanho ou densidade do disco? Seria necessário precisão e muitas tentativas e ajustes para saber onde colocar a rede de alfinetes – talvez milhões – mas isso poderia ser feito, e eventualmente, você seria capaz de usar a placa Plinko para detectar diferenças sutis entre os discos que não são facilmente detectáveis ou mensuráveis de outra forma, porque elas cairiam em uma determinada caixa que descrevia o disco. É claro que isso não seria, na verdade, prático no mundo físico, mas são exatamente o tipo de coisa para a qual computadores poderosos são bastante úteis.

Redes neurais bem treinadas superam os humanos em muitas tarefas de classificação de imagens, como leitura de endereços digitados ou manuscritos e códigos postais em cartas, um dos primeiros casos de uso de redes neurais no mundo real. Isto também é verdade no caso de determinar o liveness em imagens de selfie. Pode parecer simples, mas pode ser extremamente desafiador detectar liveness com o olho humano. Você consegue identificar as imagens falsificadas no grupo de fotos abaixo?



Consulte a página 9 para obter respostas sobre e quais são paródias.

4 The Price is Right Wiki https://priceisright.fandom.com/wiki/Plinko

Quanto mais dados verdadeiros puderem ser usados para treinamento, melhor. Mas uma segunda coisa que é útil para entender o liveness de apenas uma imagem é que a a precisão da rede neural não depende apenas do quantidade de dados de treinamento, mas em sua qualidade. Em outro palavras, se entrar lixo, vai sair lixo; treinamento com baixa qualidade de dados pode realmente degradar seu desempenho. O que determina a qualidade dos dados? Deve estar corretamente marcado. Deve ser representativo dos casos encontrados no mundo real. Deve abordar casos extremos, e deve evitar impor preconceitos nos seus resultados.

Treinar redes neurais é tanto arte quanto ciência; sempre há técnicas que os cientistas de dados podem descobrir e aplicar para otimizar ainda mais o desempenho. Usando o leitor de código postal como exemplo, o desempenho poderia ser otimizado treinando a rede para distinguir entre diferentes estilos de escrita de todo o grupo de dígitos para ajudar a informar decisões sobre quais são os números individuais. Para detecção de liveness, isso pode significar não apenas treinar a rede para determinar "ao vivo ou falsificado?", mas sim "impresso em papel?", "exibido em tela?", ou "bordas detectadas?". Para fazer essas avaliações, recursos sutis de imagem podem ser usados como entrada para as redes para informar essas decisões, como o efeito moiré[5], espectro de cores e outros e depois pesar vários resultados para determinar uma conclusão final.

"...a precisão da rede neural depende não apenas da quantidade de dados de treinamento, mas também de sua qualidade."

A ID R&D investiu muitos anos de pesquisa para determinar quais características da imagem deveriam ser examinadas e como elas deveriam ser combinadas, e está continuamente fazendo melhorias.

O software funde a saída dessas redes neurais para produzir uma pontuação de liveness e, em seguida, mapeia a pontuação para um valor de função de distribuição de probabilidade entre 0 e 1. Esse mapeamento depende da calibração, que por sua vez é baseada no uso de dados do mundo real para encontrar o valor certo. equilíbrio entre falsa aceitação e falsa rejeição para um caso de uso específico. O IDLive Face pode ser ajustado para otimizar o desempenho em diferentes ambientes usando calibração, o que não requer modificações no design da rede neural.

5 O efeito moiré é o fenômeno de padrões criados quando padrões são sobrepostos. Um exemplo é quando uma fotografia digital é tirada de uma tela digital. A resolução da câmera e da tela interferem entre si, causando um padrão de linhas na fotografia resultante.

Como a eliminação do atrito agrega novos clientes e valor

A medição do desempenho do algoritmo de liveness é análoga ao do matching biométrico. Ambos exibem erros de falsos positivos e falsos negativos com uma compensação inerente entre segurança e conveniência. O sistema pode ser configurado para otimizar qualquer um deles.

No matching, uma taxa de falsos positivos indica a frequência de correspondências incorretas entre amostras genuínas e falsas e representa uma maior ameaça à segurança. A taxa de falsos negativos aponta para rejeições de clientes genuínos que impactam negativamente a experiência do usuário.

Na detecção de liveness, APCER[6] é a taxa de erro na detecção de um ataque de apresentação, e os fornecedores de tecnologia de liveness apregoam um APCER baixo ou até próximo de zero como medida do nível de segurança que ele oferece. Mas dada a compensação inerente entre erros falso-negativos e falso-positivos, um APCER baixo pode custar um BPCER alto[7], a taxa de erro na classificação de clientes legítimos.

O atrito contribui para um BPCER mais elevado que também é menos previsível

Conforme discutido, uma abordagem de detecção de liveness"ativa" depende de interações com o usuário para ajudar a avaliar o liveness, enquanto uma abordagem "sem interação" é transparente para o usuário e normalmente usa apenas as mesmas imagens usadas para comparação biométrica. Um BPCER pode ser piorado pelo atrito introduzido por uma técnica de liveness ativa. Frustração, distração e erros na interpretação ou execução de instruções podem aumentar a frequência de interrupções e falhas e podem ser particularmente impactantes em um processo de onboarding digital, onde os usuários são novos e realizam tarefas pela primeira vez. Além disso, o atrito ao usuário introduz variáveis do comportamento humano que são difíceis de prever e medir, e por isso o BPCER observado no mundo real de uma solução ativa pode ser superior ao medido em laboratório e a diferença pode ser significativa.

Estudo de caso: o ROI da atualização da vivacidade ativa para a vivacidade passiva, de um quadro e sem atrito

A necessidade de impor inconvenientes aos usuários legítimos (um BPCER alto) para atingir uma meta de segurança (um

Processo de liveness sem interação usando apenas uma imagem

O usuário tira uma selfie

A imagem da selfie é comparada biometricamente para determinar uma correspondência

A mesma imagem selfie é usada para a verificação de liveness

O software IDLive Face usa DNNs e algoritmos proprietários para analisar a imagem em busca de liveness e falsificações

O sistema retorna um pontuação de probabilidade de liveness

6 APCER é um acrônimo para Attack Presentation Classification Error Rate.

7 BPCER é um acrônimo Bona Fide Classification Error Rate

APCER baixo) pode levar ao abandono de aplicativos e até mesmo à perda de clientes.

Um parceiro da ID R&D com um grande cliente no setor de serviços financeiros usava uma solução de detecção ativa de liveness. Tinha uma baixa taxa de APCER anunciada, mas em campo eles estavam enfrentando uma alta taxa de interrupções; um BPCER observado na faixa de 40%, o que é bastante alto e possivelmente não viável operacionalmente.

Estudo de caso: o ROI da atualização do liveness ativo para o liveness sem interação com apenas uma imagem

A necessidade de impor inconvenientes aos usuários legítimos (um BPCER alto) para atingir uma meta de segurança (um APCER baixo) pode levar ao abandono de processor e até mesmo à perda de clientes. Caso em questão, um parceiro da ID R&D com um grande cliente no setor de serviços financeiros estava usando uma solução de detecção ativa de liveness. Tinha uma baixa taxa de APCER anunciada, mas em campo eles estavam enfrentando uma alta taxa de interrupções de aplicativos; um BPCER observado na faixa de 40%, o que é bastante alto e possivelmente não viável operacionalmente.

Com apenas 60% dos clientes capazes de solicitar uma conta sem interrupção, o impacto na aquisição de clientes foi substancial. Então, eles optaram por tentar uma abordagem sem interação e sem atrito, apoiada pelo IDLive® Face. Ao contrário de uma abordagem ativa, o IDLive Face usa apenas a mesma imagem selfie usada para matching biométrico, adicionando nenhum esforço à experiência do usuário. Zero esforço adicional significa zero potencial de erro do usuário contribuído pela detecção de liveness[8].

De 60% a 95%+ das taxas de conclusão para novas solicitações

Os resultados da atualização foram dramáticos. As inscrições de novos clientes passaram de uma taxa de conclusão de 60% para mais de 95%. Isto significa que mais de um terço de todos os requerentes deixaram de ser interrompidos nas suas candidaturas e passaram a concluí-las sem interrupção. A mudança foi implementada sem degradar o desempenho da detecção de falsificação, ou seja, sem aumento no APCER.

A melhoria foi tão substancial que certamente teve não apenas um grande impacto na satisfação de uma ampla faixa da base de clientes, mas também nas finanças da empresa. Embora não seja explicitamente medido neste caso, um aumento na taxa de conclusão de mais de 50% provavelmente teve um impacto comparável na aquisição de clientes e nas receitas.

Eliminar atritos e reduzir abandonos: o impacto financeiro

Cada cliente potencial é valioso; mas aqueles que já iniciaram o processo de onboarding serão uma perda particularmente trágica. Podemos estimar o valor agregado por esses clientes que, de outra forma, teriam ido para outro lugar. Considere uma amostra de um milhão de solicitações de abertura de contas iniciados antes e depois da implementação da detecção de liveness sem ativo. Sem o atrito adicional do liveness ativo, mais de 350.000 desses clientes que sofreram interrupções no onboarding agora as concluíam sem interrupção. Mesmo que apenas metade dos candidatos interrompidos abandonem completamente as suas candidaturas, podemos estimar um aumento no valor para o banco na ordem dos 350 a 700 milhões de dólares, assumindo um valor vitalício do cliente bancário de retalho (CLV) de 2.000 a 4.000 dólares.[9] por cliente.

8 Um benefício adicional de uma abordagem passiva é que nenhuma informação é fornecida ao fraudador sobre como tentar derrotá-lo, o que pode diminuir o APCER.

9 Olhando além dos produtos para o valor vitalício do cliente, Sherief Meleis, Novantas LLC

A abordagem com apenas uma imagem é a melhor opção para detecção de liveness para melhorar a segurança enquanto evita atrito do usuário, abandonos, incerteza, e receita perdida

É geralmente entendido que com a biometria surge uma compensação inerente entre falsos negativos e falsos positivos que as partes interessadas precisam levar em consideração ao projetar um sistema. Este caso específico ilustra que, no caso da detecção de liveness, uma alta taxa de BPCER pode ter um impacto enorme nas taxas de conclusão, na satisfação do cliente e, em última análise, nos resultados financeiros. O atrito introduzido por uma abordagem ativa tenderá a resultar em um BPCER mais elevado para um determinado APCER alvo. Além disso, a imprevisibilidade do comportamento humano torna difícil extrapolar o desempenho num ambiente controlado para operações do mundo real. Com cada novo cliente bancário acrescentando milhares de dólares de valor a um banco, a diferença causada pelo atrito pode ter um grande impacto financeiro.

Respostas: As imagens 1, 3 e 5 são reais e as demais são paródias. Neste caso, a IA não apenas automatiza, mas também melhora um processo.

www.idrnd.ai