Single-Frame Facial Liveness Detection:

How It Works to Reduce User Friction and Abandonment

ID R&D
a Mitek company

The ability to ascertain facial liveness from a single-image frame enables high-performance presentation attack detection without adding friction to the user experience. A "passive" approach to liveness that is frictionless for the user eliminates the customer abandonments caused by the complications of an "active" approach. The result is fewer lost customers and more revenue. This paper discusses how single-frame facial liveness detection works and presents a case study of its impact on the onboarding operations of a retail bank.

What is passive facial liveness?

In remote identity verification and authentication processes that use biometrics, liveness detection is critical in preventing *presentation attacks*[1], or "spoofs". Common attacks include presentation of faces printed on paper or displayed on digital screens, video playbacks, and facial masks.

Approaches to liveness detection can be categorized as either *active* or *passive*. An active approach relies on interaction with the user, such as by instructing them to blink, smile, turn their head, or move their device while on camera and then detecting their reaction. The process takes more time and provides information to fraudsters that can potentially be used to defeat the security mechanism.

For example, free animation software makes it easy to create a short video playback of smiling, blinking, head turning[2]. Wearing a paper mask with eye and mouth cutouts can fool active liveness systems. Furthermore, an active approach requires multiple image frames or video, demanding more data processing and transmission, adding still more time and cost to the identity authentication process.

In contrast, a passive liveness technique requires no instruction, commands, or response from the user. Not all passive liveness is the same. Some approaches do not rely on user interaction but do use video or multiple images during the capture process, which tends to demand more processing and data transmission and thus adds friction in the form of latency and costs of computing infrastructure and networking.

The ideal passive liveness approach uses the same selfie for liveness as was already captured for face matching, eliminating the costs of processing video or multiple frames and removing the added user friction.

The ideal passive liveness approach uses the same selfie for liveness as was already captured for face matching, eliminating the costs of processing video or multiple frames and removing the added user friction.

ID R&D provides this ideal single-frame passive liveness capability via its product IDLive® Face.

[1] A presentation attack is a method of attempted fraud by which a bad actor presents a spoof instead of a live and present sample. Examples include a printed copy, an image on a digital display, or a mask.

[2] Examples include MotionPortrait, Deep Nostalgia, and Avatarify.

A summary of the advantages of single-frame liveness

Following is a summary of the advantages of single-frame liveness for both users and implementers of face authentication systems. The result is an easier, faster, more secure, and less costly solution, which adds no user friction and therefore does not contribute to user abandonment or a negative experience.

1. There is no need to educate or instruct the user, and no need for movements, gestures, or video to recognize them as a live person.

2. The same frame used for facial matching is used for liveness checking, substantially reducing the complexity of the solution design, support, and maintenance.

3. There is no requirement for changes to the user interface or communication interfaces. A back-end application simply performs one API call to IDLive Face deployed in the back-end infrastructure. This makes integration quick and straightforward for developers.

4. Minimal details are sent from the device, reducing latency. A single image is generally no larger than 300 kB when using ID R&D recommended settings as compared to solutions that send multiple frames or video. If you are already sending the image to the server, there is no added overhead for data transmission.

5. The user experience provides no information to fraudsters on how to defeat the security mechanism. There is not a separate liveness step that the fraudster can attack separately from the face matching.

When using single-frame passive liveness detection, the user is unaware that a liveness check is happening, offering a frictionless experience for users and simplified for developers:

- No instruction, command, and response for the user to follow
- The same selfie image used to for matching and for liveness
- Supports browser or native apps
- Highly simplified software integration, maintenance, and support
- Minimal device-server data interchange
- No information to help fraudsters

How single-frame liveness works

How is it possible to so accurately assess liveness from a single image, when other solutions leverage full video streams that capture user movement and interaction?

The short answer is that single-image liveness uses AI. But digging deeper, the first thing to understand about single-frame liveness is that there are features of a digital image that–while not easily visible to the naked eye–can be detected and measured using computer vision. An analogy is the difference between evaluating the picture quality of your TV at home versus at an electronics store, where many are on display and showing the same content and so the differences between TVs are easy to spot. Computer vision can detect and precisely measure these differences using math.

But how does AI work (the abridged version)? Artificial intelligence is an umbrella term that encompasses several *machine learning* techniques, many of which involve training of a *deep neural network* to perform as a *classifier*[3]. These neural networks can be thought of as an array of equations that are arranged such that outputs of each column of equations provides inputs to multiple column of equations in series (which makes them a "deep" network).

The training process involves using *ground-truth* input and output data–essentially input data that is manually tagged with "right-answer" output data–to train the network what the correct output is for a given input. The inputs and associated outputs are used in the training process to generate a set of coefficients for the equations in the neural network. Different coefficients are tried until the overall error rate is minimized, which could take millions of calculations and several iterations. It's a brute-force process of trial and error that computers are very good at. Once trained, the neural network uses a set of coefficients that minimizes error and can now classify input data it has never seen before.

[3] A classifier is a type of software designed to automatically categorize input data. Among the earliest machine learning-powered classifiers were used by the Postal Service to read the digits of handwritten zip codes.

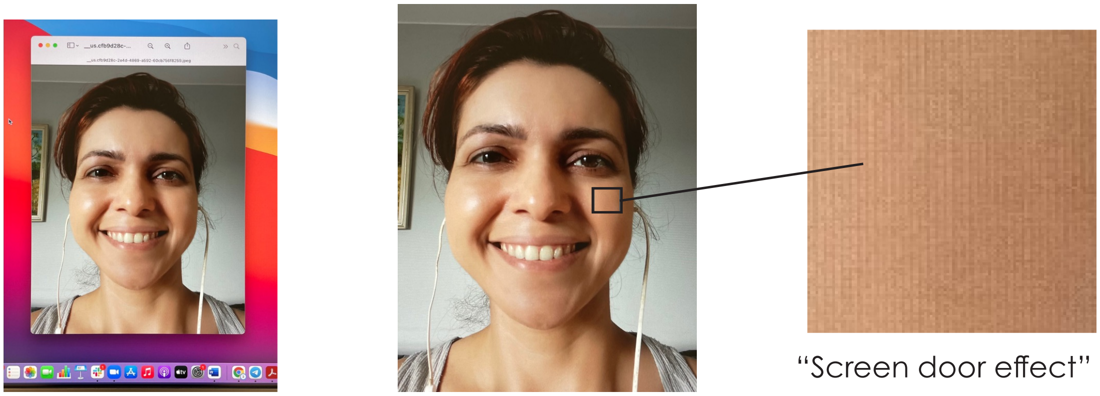
Neural Networks in Practice

Detecting “Screen Replay” Presentation Attacks

Screen replay presentation attacks are when a photo displayed on a digital screen is presented as proof of identity instead of a live selfie. While difficult to see with the naked eye, there is evidence of the attack revealed in this single image using computer vision techniques.

The “Screen Door Effect” is what happens when a photo of a digital screen is taken at high resolution. While very small, there is a grid of spaces between the pixels of the display that can be seen under magnification and detected by a high-resolution camera.

WHAT FRAUDSTERS DO | WHAT CAMERA SEES | WHAT NEURAL NETWORK SEES



“Screen door effect”

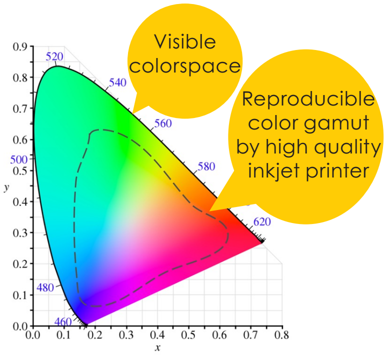
Detecting “Printed Copy” Presentation Attacks

The color gamut of a printing device is determined by the hue, saturation, and lightness of its inks (cyan, magenta, yellow, black), as well as the brightness and other characteristics of the substrate on which they are printed. Colors visible to the human eye have a larger color gamut than printing devices that use CMYK inks, especially in deep blues and blacks. Printing an image captured by a digital camera requires transforming the image from the camera’s RGB color space to the printer’s CMYK color space. During this process, the colors from the RGB which are outside the gamut must be converted to approximate values within the CMYK space gamut. For these reasons, the printed image will be less vivid than the original image captured by the camera. When a color is “out of gamut,” it cannot be faithfully converted to the target device.

WHAT FRAUDSTERS DO

WHAT CAMERA SEES

WHAT NEURAL NETWORK SEES



CIE xy chromaticity diagram

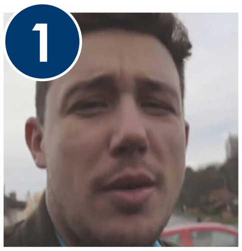
Neural networks - a physical analogy

A physical analogy of a neural network is Plinko[4], a game played on "The Price Is Right" game show for forty years, where a round disk is dropped down a vertical board full of pins; the ball hits the pins as it drops, traveling randomly and then landing into one of several numbered boxes at the bottom that have different prizes for the player.

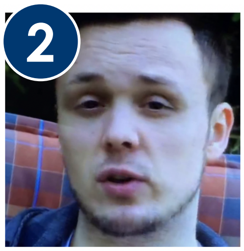
In a normal Plinko game, the disk's route and end-state is random. But what if, through trial and error, we could precisely arrange the pins such that the disk would always land in a particular box based on its physical features; say, the size or density of the disk? It would take precision and many drops and adjustments to learn where to put the network of pins–maybe millions–but it could be done, and eventually you would be able to use the Plinko board to detect subtle differences between the disks that are not easily detectable or measurable otherwise, because they would drop into a certain box that described the disk. These readjustments would of course not actually be practical in the physical world, but they are just the type of thing that powerful computers are quite useful for.

Well-trained neural networks outperform humans for lots of image classification tasks, like reading typed or handwritten addresses and postal codes on letters, one of the first real-world use cases for neural networks. This is also true in the case of determining liveness in selfie images. It may seem straightforward, but it can be extremely challenging to detect liveness with the human eye. Can you spot the spoofed images in the group of photos below?

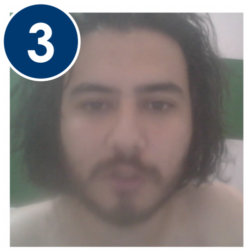
1



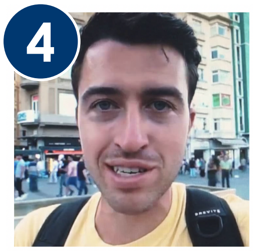
2



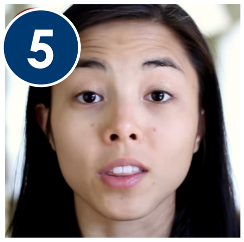
3



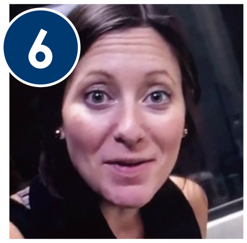
4



5



6



See the page 9 for answers to which images are live and which are spoofs.

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

The more ground-truth data that can be used for training, the better. But a second thing that's helpful towards understanding single-frame liveness is that the accuracy of the neural network relies not just on the quantity of training data, but on its *quality*. In other words, garbage in, garbage out; training with low-quality data can actually degrade its performance. What determines the quality of ground truth data? It should be correctly tagged. It should be representative of cases found in the real world. It should address corner cases, and it should avoid imposing biases on its results.

Training neural networks is art as much as science; there are always techniques data scientists can discover and apply to further optimize performance. Using the postal code reader as an example, performance could be optimized by training the network to distinguish between different handrwriting styles of the whole group of digits to help inform decisions on what the individual numbers are. For liveness detection, this might mean not just training the network to determine "live or spoof?", but rather "printed on paper?", "displayed on screen?", or "edges detected?". To make these assessments, subtle image features can be used as input to the networks to inform these decisions, such as moiré[5] effect, color spectrum, and others, and then weighing multiple results to determine a final conclusion.

...the accuracy of the neural network relies not just on the quantity of training data, but on its *quality*.

ID R&D has invested many person-years of research to determine what image features should be examined and how they should be combined, and is continuously making improvements.

Software fuses the output of these neural networks to produce a liveness score, and then maps the score to a probability distribution function value between 0 and 1. This mapping depends on calibration, which in turn is based on using real-world data to find the right balance between false acceptance and false rejection for a particular use case. IDLive Face can be tuned to optimize performance for different environments using calibration, which does not require modifications to the neural network design.

[5] Moiré effect is the phenomena of patterns created when patterns are superimposed. An example is when a digital photograph is taken of a digital screen. The resolution of the camera and of the screen interfere with each other to cause a pattern of lines on the resulting photograph.

How eliminating friction adds new customers and value

The measurement of liveness algorithm performance is analogous to that of biometric matching. Both exhibit false-positive and false-negative errors with an inherent tradeoff between security and convenience. The system can be configured to optimize for either.

In matching, a false-positive rate indicates the frequency of incorrect matches between genuine and impostor samples and represents a higher security threat. The false-negative rate points to rejections of genuine customers that negatively impact user experience.

In liveness detection, APCER[6] is the rate of error in detecting a presentation attack, and liveness technology vendors tout a low or even near-zero APCER as a measure of the level of security it affords. But given the inherent tradeoff between false-negative and false-positive errors, a low APCER can come at the cost of a high BPCER[7], the error rate in classifying bona fide customers as legitimate.

Friction contributes to a higher BPCER that is also less predictable

As discussed, an "active" liveness detection approach relies upon interactions with the user to help assess liveness, while a "passive" approach is transparent to the user, and typically uses only the same images used for biometric comparison. A BPCER can be made worse by the friction introduced by an active liveness technique. Frustration, distraction, and errors in interpreting or executing upon instructions can all increase the frequency of interruptions and failures, and can be particularly impactful in a digital onboarding process, where users are new and performing tasks for the first time. Furthermore, user friction introduces variables of human behavior that are difficult to anticipate and measure, so the BPCER observed in a real-world deployment of a high-friction solution can be higher than planned for, and the difference can be significant.

Case study: the ROI from upgrading from active liveness to passive, one-frame, frictionless liveness

A need to impose inconvenience on legitimate users (a high BPCER) in order to achieve a security target (a low APCER) can lead to abandoned applications and even lost customers. Case in point, an ID R&D partner with a large customer

Passive liveness detection process using a single-frame approach

User takes a selfie

Selfie image is biometrically compared to determine a match

The same selfie image is used for the liveness check

IDLive Face software uses DNNs and proprietary algorithms to analyze the image for liveness and for spoofs

The system returns a liveness probability score

[6] APCER is an acronym for Attack Presentation Classification Error Rate.
[7] BPCER is an acronym for Bona Fide Classification Error Rate.

in the financial services sector had been using an active liveness detection solution. It had a low advertised APCER rate, but in the field they were experiencing a high rate of application interruptions; an observed BPCER in the range of 40%, which is quite high and arguably not operationally viable.

Case study: the ROI from upgrading from active liveness to passive, one-frame, frictionless liveness

A need to impose inconvenience on legitimate users (a high BPCER) in order to achieve a security target (a low APCER) can lead to abandoned applications and even lost customers. Case in point, an ID R&D partner with a large customer in the financial services sector had been using an active liveness detection solution. It had a low advertised APCER rate, but in the field they were experiencing a high rate of application interruptions; an observed BPCER in the range of 40%, which is quite high and arguably not operationally viable.

With only 60% of customers able to apply for an account without interruption, the impact on customer acquisition was substantial. So they opted to try a passive, frictionless approach as supported by IDLive® Face. Unlike an active approach, IDLive Face uses only the same single selfie image used for biometric matching, adding zero effort to the user experience. Zero added effort means zero potential for user error contributed by liveness detection[8].

From 60% to 95%+ completion rates for new applications

The results of the upgrade were dramatic. New customer applications went from a 60% completion rate to over a 95% rate. This means that over a third of all applicants went from being interrupted in their applications to completing them without interruption. The change was implemented without degrading spoof detection performance, i.e. without an increase in the APCER.

The improvement was so substantial that it surely had not only a big impact on satisfaction across a broad swath of the customer base, but also on company financials. While not explicitly measured in this case, an increase in the rate of completions of over 50% likely had a comparable impact on customer acquisition and revenue.

Removing friction and reducing abandonments: the financial impact

Every prospective customer is valuable; but those who have already begun the onboarding process are particularly tragic to lose. We can estimate the value added by these customers who would have otherwise gone elsewhere. Consider a sample of one million account applications initiated both before and after the implementation of passive liveness detection. Without added friction from active liveness, over 350,000 of these customers who experienced interruptions to their onboarding were now completing them without interruption. Even if only half of interrupted applicants abandon their applications altogether, we can estimate an increase in value to the bank in the range of about $350 to $700 million, assuming a retail banking customer lifetime value (CLV) of $2,000-$4,000[9] per customer.

[8] An added benefit of a passive approach is that no information is provided to a fraudster on how to attempt to defeat it, which can lower the APCER.

[9] Looking Beyond Products to Customer Lifetime Value, Sherief Meleis, Novantas LLC

Single-frame is a preferred approach to liveness detection to improve security while avoiding user friction, abandonments, uncertainty, and lost revenue

It's generally understood that with biometrics come an inherent tradeoff between false negatives and false positives that stakeholders need to factor in when designing a system. This particular case illustrates that in the case of liveness detection, a high BPCER rate can have a massive impact on completion rates, customer satisfaction, and ultimately the bottom line. The friction introduced by an active approach will tend to result in a higher BPCER for a given target APCER. Furthermore, the unpredictability of human behavior makes it difficult to extrapolate performance in a controlled setting to real-world operations. With each new banking customer adding thousands of dollars of value to a bank, the difference made by friction can have a big financial impact.

Answers: Images 1, 3 and 5 are real and the others are spoofs. In this case AI not only automates but also improves a process.

ID R&D

a Mitek company

www.idrnd.ai