USENIX

THE ADVANCED COMPUTING
SYSTEMS ASSOCIATION

Red Bleed: A Pragmatic Near-Infrared Presentation Attack on Facial Biometric Authentication Systems

Bowen Hu, Kuo Wang, and Chip Hong Chang, *Nanyang Technological University*

https://www.usenix.org/conference/usenixsecurity25/presentation/hu-bowen

This paper is included in the Proceedings of the
34th USENIX Security Symposium.

August 13–15, 2025 • Seattle, WA, USA

978-1-939133-52-6

Open access to the Proceedings of the
34th USENIX Security Symposium is sponsored by USENIX.

Red Bleed: A Pragmatic Near-Infrared Presentation Attack on Facial Biometric Authentication Systems

Bowen Hu
School of Electrical and Electronic Engineering
Nanyang Technological University

Kuo Wang
School of Electrical and Electronic Engineering
Nanyang Technological University

Chip Hong Chang
School of Electrical and Electronic Engineering
Nanyang Technological University

Abstract

Facial recognition is the most prevalent biometric modality in commercial verification and identification systems, which typically operate under near-infrared (NIR) illumination. Such systems are generally considered secure on the premise that no commercial screen display can readily enable an NIR-based video presentation attack. However, this work demonstrates a critical vulnerability of NIR biometric authentication systems without advanced anti-spoofing mechanisms by a presentation attack, named Red Bleed attack, demonstrated on a widely used enterprise-grade face authentication system through a custom-built liquid crystal display (LCD) that costs less than 400 USD.

Due to the scarcity of NIR video samples, it is more feasible to sneak RGB images in the visible (VIS) spectrum through, for instance, covert secret photography, photos posted on social media or screen captures during video conferencing. Besides using live captured NIR video of the target subject's face, we also propose a novel identity-preserved NIR face generative framework that combines a Variational Autoencoder (VAE) to convert VIS images into the NIR domain for this attack. In conjunction with an advanced face swapping technique, an RGB video can be transformed into a video with NIR face, enabling a more sneaky and pragmatic 2D presentation attack on NIR face biometric authentication, demonstrated also on the COTS face authentication module.

The hardware design and source code supporting our findings are publicly available at Zenodo. This vulnerability has been reported to Microsoft and the vendors of the three evaluated COTS Windows Hello face recognition modules. The reported behavior has been confirmed by the Microsoft Security Response Center (MSRC). Microsoft has published this vulnerability under the Common Vulnerabilities and Exposures CVE-2025-26644 and released a patch in KB5055523 to address this issue in its April 2025 security update.

1 Introduction

Face recognition has become one of the most prominent biometric techniques in e-commerce, e-government, and personal devices [2, 18]. Due to the various advantages, the majority of popular facial recognition devices have unanimously chosen to operate at near-infrared (NIR) wavelengths. For instance, Windows Hello operates at 850nm. Utilizing NIR can provide consistent images under varying lighting conditions (while also allowing for subtle variations in appearance, including facial hair, makeup, etc.) [62]. The longer wavelength and lower energy of NIR light reduce the risk to human eyes. NIR is invisible to the human eye, ensuring it does not affect the user's visual experience. The technology for NIR LEDs is mature and highly efficient. To a large extent, it is widely believed that using NIR can enhance the security of facial recognition devices [62]. This belief is founded on the assumption that there is no NIR display available, thereby preventing presentation attacks. The reason for this assumption is easily understandable. As display panels are highly sophisticated products, no company would invest in developing a display for NIR video that is invisible to the human eye, which has practically no apparent use in reality.

However, this assumption no longer holds true. In this study, we demonstrate that with a cost of less than 400 USD, an ordinary liquid crystal display (LCD) can be modified into an NIR display device. LCD is a widely used display technology today. We utilized an LCD module with its backlight removed, replaced it with an 850nm NIR backlight, and added the wired-grid polarizers. This modified LCD module is capable of displaying NIR images/videos with high contrast.

We further demonstrate that this modified LCD can be used to perform a video replay attack on facial authentication systems. We chose Windows Hello as the experimental target and executed a presentation attack on this facial recognition system using the NIR display. We used an infrared camera to capture NIR facial video of the target volunteer and played it on the NIR display, which resulted in a high success rate in breaching Windows Hello facial authentication.

Due to the required specialized sensors and illumination, genuine NIR samples of target faces are not always easy to acquire stealthily. To make the attack more practical, we also explore advanced deep generative models to synthesize NIR images from visible (VIS) inputs. Thus, this study represents the first known successful 2D presentation attack on a well-known and pervasively deployed commercial facial authentication system utilizing NIR illumination and the first attempt to exploit generative techniques against such a system.

Windows Hello Facial Authentication. Windows Hello supports multiple authentication methods, including PIN, facial recognition, and fingerprint recognition [63]. Compared to traditional password-based authentication, it offers higher user acceptance and reduced authentication time. As noted by Farke et al. [24], the facial recognition component in Windows Hello is widely regarded as both secure and user-friendly.

The Windows Hello facial recognition module typically comprises an 850nm NIR light source, an NIR camera, and a conventional RGB camera. By default, only the NIR camera is utilized for facial recognition. However, enabling the enhanced anti-spoofing (EAS) feature activates both the NIR and RGB cameras. With EAS enabled, the proposed attack method **fails** to bypass the Windows Hello authentication. While this configuration can improve recognition security, it also prolongs authentication time and is less robust under low-light conditions because it necessitates high-quality VIS images. Also, it loses the convenience of using the privacy shutter on the VIS camera without interfering with the NIR system. In this work, we perform the Windows Hello verification process under the default (**NIR-only**) setting.

Overview of the attack. In this work, we investigate a critical vulnerability in the default Windows Hello facial authentication system using a modified NIR LCD. Our approach consists of two primary components. First, on the hardware side, we construct a modified LCD module using commercially available materials, substituting the standard polarizer with one that is capable of polarizing light at 850nm. This customized LCD is then shown to successfully replay authentic NIR video samples captured via an NIR camera, thereby circumventing face verification with a high success rate. We validate our method on faces registered by 22 volunteers in three different brands of Windows Hello modules, each exhibiting susceptibility to attack.

Second, from a software perspective, we develop a deep generative model, using the variational autoencoder (VAE) designed to synthesize NIR video samples from VIS images. This method addresses the difficulty of acquiring genuine NIR samples since it requires an NIR sensor and a light source to obtain a good quality live NIR video sample of the target's face for the attack. Such a generative model preserves the subject's identity attributes while converting VIS samples into the NIR domain. We further employ a publicly available face-swapping framework, DeepFaceLab [69], to substitute the synthesized NIR face into the red channel of the original VIS image. Our findings confirm that a Windows Hello–enabled system can be unlocked solely by feeding these generated NIR samples through our improvised NIR LCD. Due to the limited availability of NIR-VIS face image pairs in the public domain, our generative attack requires at least one NIR-VIS image pair of the target subject to synthesize the attack video.

We designate the vulnerability of NIR biometric authentication systems to NIR presentation attacks as **Red Bleed**, and the corresponding attacks targeting this vulnerability are referred to as **Red Bleed Attacks**. *Backlight Bleed* is a unique phenomenon in LCD where light leaks from the edges of the screen. By replacing the backlight with 850nm infrared, the name *Red Bleed* metaphorically suggests the hidden security loophole in NIR-based biometric authentication systems that is exploited by our modified LCD for this attack.

Our contribution. Our main contributions are summarized as follows:

- We propose a method to modify an LCD at low cost to enable it to display NIR images. This negates the long-standing assumption that NIR display devices do not exist.
- We demonstrate that the modified NIR display can be used to perform presentation attack. Its face video playback can successfully unlock Windows Hello facial authentication with a full success rate on all our test subjects.
- We propose a new VAE-based deep generative model to transform the VIS image samples into NIR image samples. The generated NIR video from the few-shot image learning could unlock the Windows Hello system with our assembled NIR LCD with a very high success rate of 97.93%.

2 Related Work

In this section, we review the literature on physical attacks on face biometric systems, current face presentation attack detection algorithms, and recent development of deep generative models for human face.

2.1 Attack on the Face Biometric System

Physical attacks aim to malfunction or impersonate the face recognition system, exploiting algorithmic vulnerabilities to manipulate recognition outcomes. Traditional hand-crafted recognition algorithms are particularly susceptible to variations in lighting, facial orientation, and image quality, which often leads to misclassifications under suboptimal conditions. While advances in artificial neural networks have significantly improved recognition accuracy, these networks remain vulnerable to minor input variations, which can cause drastic

changes in inference results [86]. This inherent instability exposes critical vulnerabilities, prompting widespread research and discussion.

Sharif et al. [82] demonstrated an impersonation attack by printing specific patterns on eyeglasses to deceive recognition algorithms. Similar attacks have been proposed with various physical artifacts, such as fake eyes [71] and specialized hats [45]. Adversarial makeup [99] and adversarial stickers [31] have also been designed to degrade the face biometric system's recognition performance. Zhou et al. [109] introduced a projector-based attack to cast targeted perturbations, while Shen et al. [83] projected images directly onto an adversary's face to override recognition systems. These attacks aim to mislead face recognition models into making incorrect decisions but often disregard the anti-spoofing module's capabilities. Consequently, it remains dubious that the unattested artifacts and perturbations used in such attacks can actually evade presentation attack detection mechanisms.

Moreover, these adversarial approaches typically exploit algorithm-specific weaknesses, which require access to white-box models or training data to craft robust adversarial patterns. This dependency poses significant challenges when attempting to bypass general-purpose black-box commercial face authentication systems, leaving their real-world applicability severely limited. Considering a black-box model, the Hua-pi attack [94] employs a display device to present the unaltered facial content of an authorized user to the face authentication system, which successfully bypasses the anti-spoofing module by tricking it into accepting the presentation as genuine. This attack combines multi-modality information including RGB, NIR and depth. However, the model employs laser-printed NIR photos instead of NIR videos to perform the attack; therefore, they cannot provide any temporal information in the NIR domain and fail to spoof robust NIR face biometric systems like Windows Hello face authentication.

2.2 Face Presentation Attack Detection

The impersonation of identity—i.e., spoofing a face recognition system to accept a targeted identity—is the primary concern of this work. Face presentation attack detection or anti-spoofing has received significant attention to ensure the security of face recognition systems. Common attack methods include paper attacks, video replay attacks, and 3D mask attacks [102]. Early face anti-spoofing (FAS) research primarily relied on handcrafted features such as local binary patterns (LBP) [1], histogram of oriented gradients (HOG) [15], image quality indicators [28], optical flow motion [8], and remote photoplethysmography (rPPG) clues [52] to detect these attacks. These features have been proven to be discriminative for distinguishing bona fide samples from presentation attacks, and have been integrated into various hybrid models [25,49,50,75,101]. Additionally, some early research uses CNN-based feature learning [11,30,58] for texture classification or incorporates long short-term memory to identify the attack from the temporal information [17,64,93].

Deep models directly supervised by binary classification often learn superficial cues (e.g., screen bezels) instead of more intrinsic features. In contrast, pixel-wise supervision provides fine-grained, context-specific clues, leading to more robust feature learning. Accordingly, auxiliary supervision signals such as pseudo depth maps [4,54], binary mask labels [29,56,85], and reflection maps [43,100] have been proposed to capture local live/spoof cues. In parallel, generative approaches incorporating pixel-level supervision [55,56] have recently shown promise for estimating generic spoof patterns. Although these approaches achieve robust results in within-dataset scenarios, performance often degrades on unseen domains due to overfitting. As a result, domain adaptation techniques have been introduced to FAS research [36,48,57,107].

However, these existing FAS studies focus primarily on the RGB-visible domain. Our newly proposed NIR replay attack falls outside this conventional scope, as 3D mask attacks—while also outside standard RGB—are typically considered less critical due to material-driven differences in facial texture [102]. Additionally, our NIR attack leverages video replay, which incorporates temporal information. This novel type of spoofing warrants further research to effectively detect and counteract it.

2.3 Face Generative Model

The field of facial image generation has made remarkable advancements in recent years, particularly with the development of style-based generative adversarial networks (GANs) [40–42, 61], achieving impressive results even in view-consistent synthesis [3, 9, 10, 81]. While earlier attempts focused on conditioning GANs using modalities like text [37,38], the advent of diffusion models [19,32,84,96] has propelled the field to new heights. Models such as DALLE-2 [73], Imagen [79], and Stable Diffusion [76] have set benchmarks in creative image generation by leveraging large-scale datasets. Subsequent models have further enhanced detail and realism [70]. Beyond traditional approaches, recent developments have enabled 3D human generation through general [44,67] and text-guided diffusion models [35,104]. For greater control over generative outputs, image-based conditioning alongside text has been introduced. Techniques like ILVR [14] use iterative refinement with target images, while SDEdit [60] adds noise to the input.

In the domain of subject-conditioned generation, face recognition models play a pivotal role in extracting identity features from facial images. These models generate facial embeddings to measure identity similarity [7,18], enabling their inversion to produce facial images from identity embeddings in black-box settings [59, 74, 97]. These approaches use GAN and diffusion architectures for zero-shot generation [21,39]. However, current inversion methods often rely

on low-resolution datasets [21, 87] or high-quality but limited images [39], which have severely restricted their generalizability.

Recent advancements in Stable Diffusion [76] have introduced innovative paradigms like Textual Inversion [26] and DreamBooth [77], which fine-tune diffusion models using subject-specific identifiers to replicate individual subjects. Enhancements like HyperDreamBooth [78] use LoRA [34] and hypernetworks to optimize a text-to-image diffusion model using only a single input image. Encoder-based approaches such as E4T [27] and ProFusion [108] also contribute to faster optimization, while CustomDiffusion [47] selectively tunes network parameters. Other approaches, like Celeb-Basis [103] and StableIdentity [90], use celebrity embeddings to condition text-based models, while Kosmos-G [65], a multi-modal perception model, accommodates diverse inputs, including facial images.

More recently, direct conditioning of diffusion models on facial features has been explored for personalization without tuning. Techniques like PhotoVerse [12], and PhotoMaker [53] employ CLIP [72] image encoders to represent subjects, though they are limited by CLIP's encoding capabilities. Approaches such as Face0 [89], DreamIdentity [13], and PortraitBooth [68] enhance fidelity by incorporating face embeddings. Notably, IPAdapter [98] introduces a decoupled cross-attention mechanism to separate subject conditioning. InstantID [91] builds upon this with a stronger ID guidance and facial landmark conditioning. FaceStudio [95] combines CLIP and ID embeddings for stylized outputs, while DiffusionRig [20] incorporates 3DMM rendering for explicit control over pose, illumination, and expression.

ControlNet [105] stands out by achieving precise spatial control in text-to-image models through trainable network components, enabling fine-tuned spatial manipulation. Meanwhile, universal guidance [6] bypasses retraining during conditioning. Although these methods combine text and spatial control to manipulate input photos, they fall short of consistently generating an individual's identity under varying conditions—a challenge effectively addressed by subject-driven generative models. Arc2Face [66] uses the Arcface [18] features as guidance to generate high-quality face image samples that preserve identity information; however, such samples are generated in the visible RGB domain, and the Arcface identity guidance is insufficient for the generation of NIR samples to spoof the Windows Hello facial authentication system.

3 Red Bleed Attack

For a long time, NIR-based biometric recognition/identification, such as iris [16] or face recognition/identification has been considered highly secure. This is largely based on the assumption that there is no display device capable of presenting images in the NIR wavelength. In this work, we demonstrate that this assumption is no longer valid. We present a novel presentation attack, called the Red Bleed attack, that can be used to spoof NIR-based biometric recognition/identification systems. We constructed an infrared display device using commercially available off-the-shelf products at a cost of less than $400. Experimental results on a diverse group of 22 volunteers representing different races, genders, and age groups show that the Red Bleed attack can successfully spoof one of the most widely used face recognition systems in the commercial market, Windows Hello, with a 100% success rate. Furthermore, we tested modules from three different suppliers, which are Dell, HP, and Lenovo, and found that all of them are vulnerable to the Red Bleed attack. The vulnerability of NIR-based biometric recognition/identification systems to the Red Bleed attack is general and not limited to a specific device or supplier.

3.1 Threat Model

To better define the scope of the attack, this section will briefly outline the threat model.

Threat Objectives: To illicitly authenticate someone by presenting a captured or forged NIR video of a legitimate enrolled user to the NIR image sensor, or conduct unauthorized access to devices or facilities, data theft, or disruption of secure services.

Security Assumptions: The implementation of the targeted face recognition system assumes that a display device in NIR does not exist. To maintain a good performance in low-light environments, only infrared image sensors are used for recognition in a wide variety of environments. The targeted face recognition system lacks advanced anti-spoofing mechanisms, such as structured light or time-of-flight sensors.

Adversary Profiling: The adversary needs to build an NIR display device using off-the-shelf components and needs to get access to the victim's NIR videos using the existing facial recognition module or NIR cameras. The adversary has direct physical access to the target device to conduct the presentation attack.

3.2 Structure of the LCD

Before we describe the Red Bleed attack, we first introduce the structure of the LCD, which is the key component of the infrared display device.

The Thin Film Transistor Liquid Crystal Display (TFT-LCD) technology has been widely used in various display devices, such as smartphones, tablets, and laptops. Since the early 1970s, LCD has gradually emerged as the predominant display technology [80].

Liquid crystals possess a property known as electro-optical effects. Under the influence of an electric voltage, liquid-crystal cells can alter the polarization state of light passing

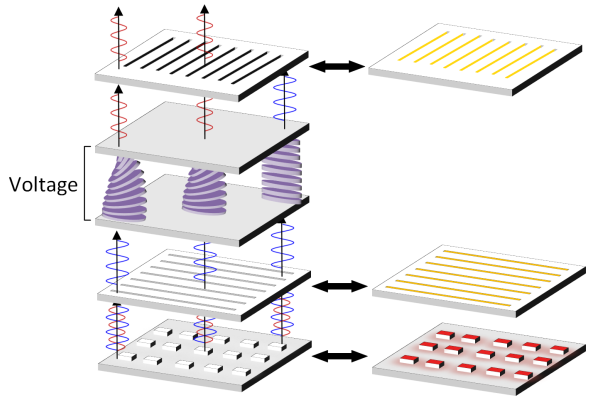


Figure 1: The structure of the LCD and our modification. In the modified design, the backlight source is changed from visible white light to NIR light, and traditional polarizers are replaced by wired-grid polarizers.

through them, a property that is crucial for display functionality. A complete LCD comprises several primary layered structures, as shown in Figure 1. From back to front, these primary layers are the *Backlight layer*, *Back Polarizer layer*, *Liquid Crystal layer*, *Color Filter layer*, and *Front Polarizer layer*. Since liquid crystal is a passive element and does not emit light by itself, it must be illuminated by a backlight layer. The backlight layer typically emits white light, which includes red, green, and blue wavelength components. The light emitted by the backlight layer passes through the back polarizer layer, which only allows light with a specific polarization state to pass through. This specific polarization direction is called transmission axis of the polarizer. The transmission axis of the back polarizer layer is perpendicular to the transmission axis of the front polarizer layer. In the absence of the liquid crystal layer, the light passing through the back polarizer layer would be blocked by the front polarizer layer, due to their perpendicular transmission axes. However, each pixel in the liquid crystal layer can change the polarization direction of light passing through it by applying an electric voltage. This change in polarization direction allows the light to pass through the front polarizer layer and be visible to the user. Different polarization directions result in varying intensities of light transmission, with light parallel to the transmission axis of the front polarizer being fully transmitted, and light perpendicular to it being completely blocked. The intensity related with the polarization direction follows the Malus's Law, as shown in Equation (1).

$$I = I_0 \cos^2(\theta) \tag{1}$$

where I and I_0 are the transmitted and incident intensities, respectively, and θ is the angle of incidence to the direction of polarization.

By applying different voltages to the liquid crystal layer, the intensity of pixels can be adjusted to display various images on the screen. For color display, the color filter layer is used to filter the white light into red, green, and blue components, which are then combined to form the desired color.

3.3 Modifying the LCD for NIR Display

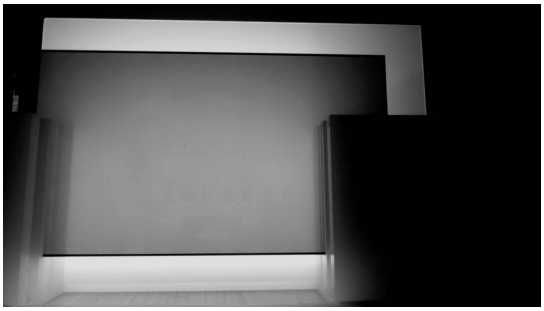
A straightforward idea is to replace the backlight of an LCD with an NIR light source to convert a standard RGB display into an NIR display. However, through practical experimentation, we discovered that simply substituting the backlight with an infrared one cannot achieve the desired effect. Figure 2a shows an image captured by an 850nm NIR camera of a scene displayed under white backlighting. The image is completely black, because the white light emitted by the backlight is not visible to the NIR camera. Figure 2b shows an image taken with the same NIR camera after replacing the backlight with 850nm NIR LEDs, where the NIR light completely penetrates the entire screen, yet no image is clearly visible.

Through extensive testing, we found that after replacing the ordinary white backlight LEDs with an 850nm NIR LEDs, the light transmitted through the display is not perfectly linearly polarized after passing through the front polarizer layer, as would be expected. Therefore, we suspect that the issue is not originated from the liquid crystal itself, but the polarizers.

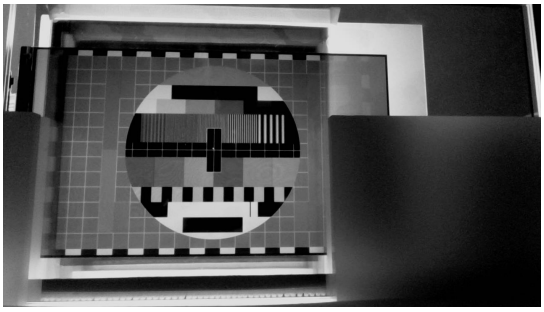
The polarizers used in liquid crystal panels are typically inexpensive and easily mass-produced polyvinyl alcohol (PVA) polarizers. These polarizers are manufactured by stretching PVA and then dyeing it with iodine [23]. The finished polarizers absorb polarized light perpendicular to their transmission axis. However, this property of absorbing polarized light does not universally apply across all wavelength ranges. For example, the *XP42 polarizer*, according to its datasheet, experiences a rapid degradation in performance for wavelengths beyond 700nm [22]. This will result in inadequately polarized light when the wavelength of the light passing through the polarizer exceeds 700nm. To verify if the problem lies with the polarizer's performance, we sought various types of polarizers that can function at NIR wavelengths. Among these, the wire grid polarizer offers excellent polarization performance across a broad wavelength range of 400-1200nm. Our experiment confirmed this, as shown in Figure 2c, where the addition of wire grid polarizers to the front and back of the liquid crystal screen allowed the LCD with 850nm NIR backlight to clearly display images.

To construct an NIR display capable of showing NIR images, we need a display panel as the core to start with. Typically, the backlight of a screen is sold encapsulated with other components, and the backlight is usually mounted on a metal frame, making the disassembly process really challenging. Removing the metal frame would leave the fragile glass substrate of the panel without support. Additionally, the backlight may be glued to other layers, making it difficult to perfectly remove a display module's backlight without damaging its other parts. Fortunately, thanks to the special demand of the

(a)



(b)



(c)

Figure 2: (a) Image captured by an 850nm NIR camera of a scene displayed under white backlighting; (b) Image taken with the same NIR camera after replacing the backlight with 850nm NIR LEDs; (c) Image taken with the same NIR camera after replacing the backlight with 850nm NIR LEDs and applying the wire grid polarizer.

masked Stereolithography (mSLA) 3D printing community, which requires ultraviolet (UV) light sources for curing resin, LCD modules designed for UV curing 3D printers with the backlight already removed can be purchased off the shelf. We selected the Sharp LS055R1SX04A model display module, which is a 5.5-inch color LCD panel with a resolution of 1440(RGB)×2560, for the next step of modification.

For the polarizer, we selected the HC12N from Asahi Kasei, a wire grid polarizer capable of providing a 1600:1 extinction ratio at the 850nm NIR wavelength. Its original size is 240mm×80mm, and we used a professional trimmer to cut this polarizer into two pieces, measuring 160mm×80mm

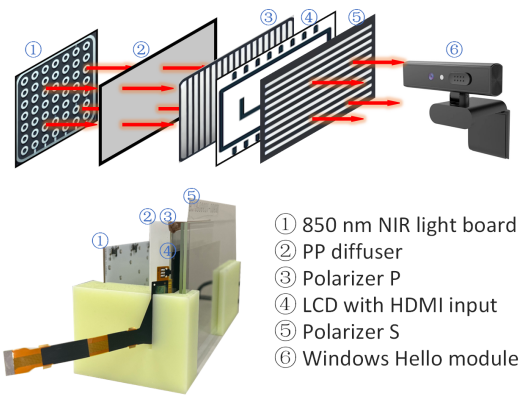


Figure 3: Our self-constructed NIR display.

and 80mm×80mm, respectively. After attempting and damaging two display panels, we chose not to remove the original polarizers from the LCD panel, despite it could potentially increase the backlight efficiency. Instead, the sliced wire grid polarizers were attached to the front and back of the display panel, with their transmission axes aligned with the original polarizer.

Additionally, we design a custom printed circuit board (PCB) to evenly mount 21 pieces of 850nm NIR LEDs to provide a sufficiently bright NIR backlight. These LEDs were arranged in a 7-series-3-parallel configuration, driven by three sets of constant current circuits. Even then, the light emitted by the circuit was not uniform enough. After trying various materials and thicknesses, a 1mm thick diffuser plate made of Polypropylene (PP) material was inserted between the LED backlight and the LCD module. This combination of the LED backlight and the PP diffuser board was able to provide the LCD module with sufficiently bright and uniform backlighting. Furthermore, our tests revealed that the NIR light could effectively and uniformly penetrate through the color filter layer of this RGB LCD panel. This means that using our externally powered NIR backlight, no compensation is needed for the slight loss in brightness.

Finally, we 3D printed a bracket to fasten these components. The final assembly is shown in Figure 3. All the components of the constructed NIR LCD are COTS material with a total cost less than 400 USD. The detailed bill-of-material (BOM) is shown in Table 1.

3.4 Red Bleed Attack on Windows Hello Face Authentication

We used this improvised 850nm NIR display to attack the Windows Hello face authentication module. Our first step was to capture a live NIR video of the target subject. We used the OmniVision OV9281 monochrome camera module paired

Table 1: The BOM of the constructed NIR LCD.

Product Name	Quantity	Price (USD)
Asaki KASEI Wired Grid Polarizer WGN 240mm×80mm	1	309.90
Sharp 5.5-inch 2K LCD LS055R1SX04A	1	30.90
PP Diffuser 150mm×100mm	1	0.07
Customized 850nm LED Board	1	15.50

Table 2: The specifications of three different brands of Windows Hello modules.

Brand	Model	Resolution	Frame rates	NIR light blink frequency
HP	830214-1U0	340×340	30	15
LENOVO	Thinkpad X1 2019 integrated	640×360	15	7.5
DELL	CN-0NVH0J	340×340	30	15

with an f2.8mm M12 lens equipped with an 850nm IR filter for this purpose.

As mentioned previously, the algorithm used by Windows Hello for face recognition is a black box to us. We do not know exactly how it is implemented. Based on our experiments with the HP module, we believe that Windows Hello uses only the NIR camera for face recognition and not the RGB camera. This is because after covering the RGB camera on the module with aluminum foil tape, Windows Hello was still able to perform correct facial recognition authentication. This behavior is expected, as these devices often need to function in poorly lit indoor environments or at night, making it a highly available design decision to rely on the NIR LEDs' illumination to obtain a clear image of the user's face. Additionally, we found that Windows Hello has no special requirements for the position or shape of the 850nm fill light source. Covering one of the 850nm LEDs with aluminum foil tape did not affect the recognition rate of Windows Hello. Moreover, even when all the 850nm LEDs were covered, Windows Hello was still able to pass facial recognition authentication with an external NIR light source. This implies that the NIR illumination requirement for capturing the live infrared facial image of a subject for authentication is not very strict.

During our initial experiment, we used the HP Windows Hello camera module to capture the NIR video of the target subject's face under the 850nm NIR illumination. We then played this video on our constructed NIR display. The Sharp LS055R1SX04A LCD panel we used is a 5.5-inch display, which is significantly smaller than the actual size of a human face. The primary reason for choosing such a small display is due to the default 80mm width of the Asahi Kasei wire grid polarizer available on the market. Fortunately for us (or unfortunately for the Windows Hello module), the lens of the NIR camera on the Windows Hello module has a shorter focal length, and a lower resolution, which allows the NIR camera of Windows Hello to stay focused even at very close distances.

According to the pinhole camera model, we can simulate the actual size of the human face by using a smaller display panel and placing it closer to the NIR camera, with the Windows Hello module aligned with the display screen at the correct distance and angle. As illustrated in Figure 4, the Windows Hello module is positioned with a calibrated geometric alignment relative to the LCD screen.

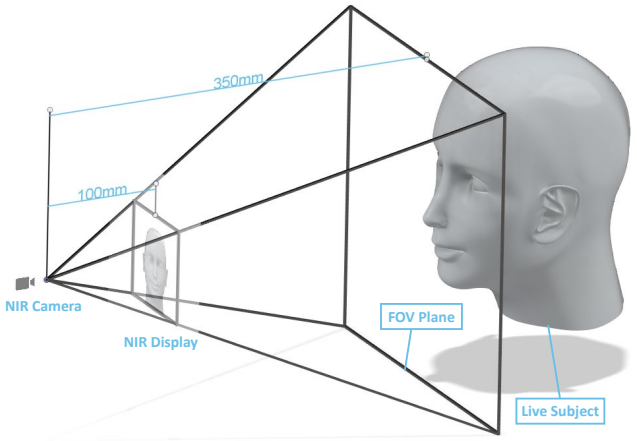


Figure 4: Illustration of the relative geometric alignment between the camera, LCD, and real face.

By capturing the actual infrared video of the test subject using the HP Windows Hello camera module, and playing this video on our constructed NIR display, we can indeed sometimes pass the facial recognition authentication of the Windows Hello module. However, the success rate is not very high, and the recognition rate is not stable. By comparing the directly captured NIR video with the NIR video that was played back on the NIR display and then recaptured by the infrared camera, we conjecture that the main issue lies in the inconsistency of brightness introduced by the missing nonlinear mapping of gamma correction in the image signal processing pipeline of the HP Windows NIR camera module, since it does not have to feed the captured image to the display

device. Through careful design of experiments, we found that after applying a certain degree of gamma adjustment to the HP module's captured NIR video and then playing it through the NIR display, it was possible to pass the Windows Hello facial recognition authentication with a 100% success rate. The gamma adjustment is shown in Equation (2), where I_{in} and I_{out} are the input and output intensities of each pixel, respectively, and γ is the gamma value.

$$I_{\text{out}} = I_{\text{in}}^{1/\gamma} \tag{2}$$

Our earlier experiments also showed that using the HP Windows Hello camera module to capture the target's live face for our Red Bleed replay attack on the higher resolution Lenovo Windows Hello module will reduce the attack success rate to 55%. For this reason, we used a higher resolution NIR camera for the experiments presented in Section 3.5.

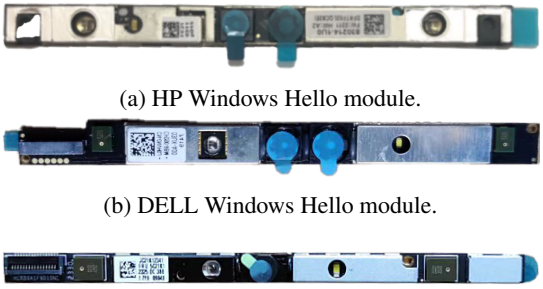
Additionally, we tested the scenario of displaying a single static NIR photo of the target and discovered that the static NIR photo of the target could not effectively unlock Windows Hello face authentication. We believe that Microsoft has successfully enhanced the anti-spoofing strength of Windows Hello by incorporating temporal information in response to some previously demonstrated spoofing cases [88]. This enhancement made static photo presentation attacks ineffective in unlocking the device.

3.5 Results of the Red Bleed Attack Using Live Captured Samples

We conducted experiments using a live NIR video captured with an OmniVision OV9281 monochrome camera module to assess the effectiveness of our assembled NIR LCD. The experimental setup and results are presented in this section.

Experiment Setting. We tested three different Windows Hello modules shown in Figure 5 in this experiment. Their models and configurations are listed in Table 2. The target laptop was a Lenovo Thinkpad X1 2019 running Windows 11 Pro Version 23H2 with the default Windows Hello setting. We captured a 6.6-second (200 frames), 1280×720 NIR video with the OmniVision OV9281 monochrome camera module, then looped it from end to start to create a 13.2-second continuous replay video for the attack. The video uses FFV1 lossless encoding. One sample attack video is provided in the supplementary file [5]. Figure 6 shows the captured real face and the replay. We recruited 22 volunteers to register their real faces in the Windows system and captured their live video samples to carry out the attacks. The participants included both male and female individuals from diverse origins of China, Singapore, India, Myanmar, Indonesia, Turkey, and the United States, with ages ranging from 20s to 60s. After setting up the devices, we conducted 20 tests per subject on each module and recorded the corresponding success rates.

Evaluation. The success rates are presented in Table 3. Three video demonstrations of successfully unlocking the three tested Windows Hello modules are provided in the supplementary file [5]. Previously, when we conducted the test on one of the subjects using the HP Windows Hello camera module to capture the 340×340 video streams, it gave a lower attack success rate of 55% on the higher resolution Lenovo device. After changing to use OV9281 camera module with ten times higher resolution and gamma adjustment in its image processing pipeline to acquire the videos, the Red Bleed attack can achieve 100% success rate on all the test subjects.



(a) HP Windows Hello module.

(b) DELL Windows Hello module.

(c) LENOVO Windows Hello module.

Figure 5: Windows Hello modules used in our experiments.

Table 3: Attack success rates of Red Breed on three different COTS Windows Hello modules.

Brand	Successful rate
HP	100%
LENOVO	100%
DELL	100%

(a) (b)

Figure 6: NIR image samples captured by the HP NIR camera. (a) The real face sample used for the attack. (b) The recaptured video frame shown on the NIR LCD (with a video player progress bar below).

4 Generation of the NIR Video

One requisite of the Red Bleed attack is the acquisition of a segment of the NIR video as short as 2∼3 seconds of the

target person under well-illuminated 850nm NIR lighting conditions. Although our tests have shown that the attack does not impose stringent requirements on lighting and equipment for capturing such a video, and it is entirely possible to use a quality NIR video captured with a very low-resolution camera module, it is still not easy to physically sneak such a video of a subject's frontal face. In comparison, the RGB video does not require any special camera module to capture. Moreover, there are diverse channels and ample opportunities to acquire a snapshot of a subject's frontal face without having to physically capture it. Also, these video clips can be easily and remotely accessed or downloaded from selfies on social media or from online video conferences.

4.1 Lessons Learned from Unsuccessful Attempts

This task turns out to be very challenging and unprecedented, as there is no demand for identity-preserved generative NIR facial video before our Red Bleed attack is introduced. The closest 2D presentation attack is the laser-printed NIR image [94] but it uses ground-truth NIR photos instead of generative images. We explored several generative methods to convert RGB images into NIR images. Some of the unsuccessful attempts are presented here to shed light on the unique problem we encountered with these seemingly plausible methods. Due to the lack of prior art, limited time and resources, there might be tricks and optimizations that we have overlooked but could potentially be applied to these baseline methods to achieve the desired outcomes. Nevertheless, lessons learned from these failed attempts have helped us understand the problem better and paved the way to a workable solution introduced in the later section.

The most straightforward method is to directly convert the VIS image to the NIR image using a simple Auto Encoder (AE) model. However, the prerequisite for this approach is the availability of a sufficiently precise VIS-NIR face dataset with pixel correspondence. Unfortunately, the highest quality dataset we could find is the CASIA NIR-VIS 2.0 face database [51], which contains 725 subjects. Each identity in the dataset has a dozen VIS and NIR photos. These photos are taken under different lighting conditions, using different cameras, at different moments; hence there is no pixel correspondence between the VIS and NIR photos. We attempted to align the photos using a facial landmark model and experimented with various possible local constraints, such as using local neural networks to try matching the pair of images with the closest landmarks, and training with losses like structure similarity index measure (SSIM) [92] or HoG [15] that are less sensitive to pixel discrepancies. However, the results were not satisfactory. The final NIR images generated by the model are very blurry and pixelated, which do not resemble the real NIR images, and fail to unlock Windows Hello with our Red Bleed gadget.

Another method we attempted was to use a GAN to generate NIR images. We used a very intuitive and mature method, CycleGAN [110], which is an unsupervised image translation GAN that can achieve conversion between two different domains without the need for paired training data. This feature aligns perfectly with our requirements. Despite its promising premise, the samples generated by CycleGAN have noticeable flaws, which prevent them from passing facial verification. In our attempts, we found that GANs are difficult to converge. Therefore, we did not delve deeper into exploring more possible GAN-based methods that require experience to overcome method-specific mode and discriminator collapse issues in model training.

Another promising candidate is the diffusion generative model. Diffusion models have shown great capabilities in generating high-quality images. There are some open-source implementations of diffusion models with pretrained weights. We tried to use Stable Diffusion v1.5 [76] for image generation, but it was evident that the model was not trained on the NIR domain. We experimented with IP-Adaper [98], an image prompt method that can adapt the diffusion model with the reference image. We also tried Textual Inversion [26], a method that can fine-tune the diffusion model with text prompts. However, both methods failed to make the Stable Diffusion model generate high-quality NIR images. We believe that the main reason for the failure is that the Stable Diffusion model was trained in the RGB domain. For uncommon data like NIR, the use of pre-trained generative models or with simple domain adaptation may not yield satisfactory outcomes.

Next, we tried training the diffusion model from scratch. The Conditional Diffusion model [33] and Guided Diffusion model [19] are two diffusion methods that we attempted. However, neither was able to produce effective generated samples. We believe that the main issue lies in the small training dataset, making it difficult to train an effective conditional diffusion image generative model from scratch to model the complex data distribution with limited data conditions. As for the Guided Diffusion model, we require a pre-trained guidance model. ArcFace [18] is a widely used face recognition model that we adopted as the guidance model. However, ArcFace is also a model trained on RGB images, and our tests showed that its performance on NIR data significantly deviated from that on RGB images. We attempted to fine-tune ArcFace to reduce the performance deviation on NIR data, but faced the bottleneck with limited dataset. Consequently, this diffusion-based approach also failed to yield high-quality generation results that can successfully unlock Windows Hello with Red Bleed.

4.2 Proposed Generative Method

Drawing from the experiences gained through the failures of various attempts, we propose a VAE-based NIR face image generation method. VAE, like GAN and Diffusion, is a gener-

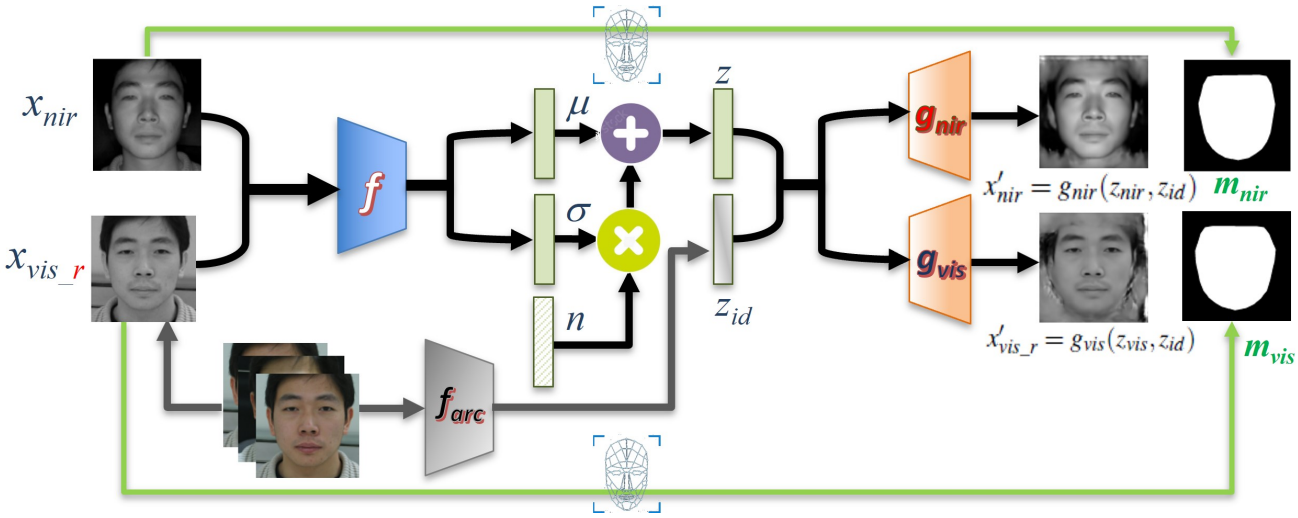


Figure 7: Block Diagram of the Proposed VIS-to-NIR Identity-preserving Face Generative Model

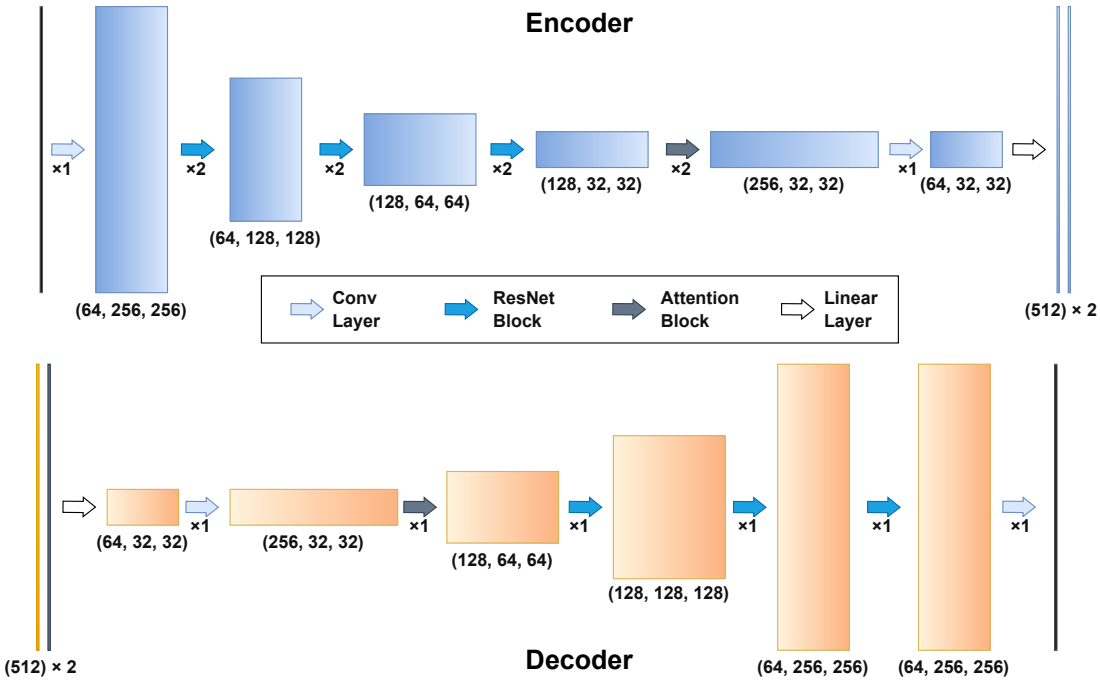


Figure 8: Structures of the encoder and decoder

ative algorithm. Although it is very similar in model structure to AE, there are considerable theoretical differences between VAE and AE. One main difference is that VAE introduces a probabilistic model into the latent space.

The block diagram of Figure 7 illustrates the proposed method. Firstly, all VIS or NIR photos containing faces are aligned and resized into square images through affine transformation. Ordinary VIS images have three channels: Red, Green, and Blue. Since we want the encoder to be compatible with NIR, which has only one channel, only the red channel of the VIS photo is input into the encoder. The encoder f will map the image x, whether NIR or VIS, to a mean μ and a standard deviation σ, as shown in Equations (3) and (4).

$$(\mu_{nir}, \sigma_{nir}) = f(x_{nir}) \tag{3}$$

$$(\mu_{vis}, \sigma_{vis}) = f(x_{vis\_r}) \tag{4}$$

Then, the respective latent output z is obtained by sampling a Gaussian distribution $\mathcal{N}$ with the corresponding mean and variance outputs of the encoder f, as expressed in Equations (5) and (6).

$$z_{nir} \sim \mathcal{N}(\mu_{nir}, \sigma^2_{nir}) \tag{5}$$

$$z_{vis} \sim \mathcal{N}(\mu_{vis}, \sigma^2_{vis}) \tag{6}$$

Additionally, we have incorporated ArcFace f_{arc} as the identity information, the embedding is denoted as z_{id} in Figure 7. Since the ArcFace model is trained in the RGB domain, we only use VIS photos to encode the target's identity, and for each identity, we calculate the average of their id embeddings, as shown in Equation (7).

$$z_{id} = \frac{1}{N}\sum_{i=1}^{N} f_{arc}(x_{vis_i}) \tag{7}$$

For the decoder, its inputs are the latent vectors z and z_{id}. Unlike the encoder, we employ two different decoders corresponding to the NIR and VIS image domains, respectively. These two decoders are represented by g_{nir} and g_{vis}, in Equations (8) and (9), respectively.

$$x'_{nir} = g_{nir}(z_{nir}, z_{id}) \tag{8}$$

$$x'_{vis\_r} = g_{vis}(z_{vis}, z_{id}) \tag{9}$$

The structures of the encoder f and the decoder g are similar to that of a 2D U-Net used in Stable Diffusion [76] but without the intermediate skip connections. Figure 8 illustrates the structures of the encoder and decoder. The encoder f consists of four levels of functional blocks, with the first three being ResNet blocks and the last being an Attention block. The output is mapped to a single channel by a 1×1 convolution layer, then flattened into one dimension, and finally passed through a linear layer to output vectors of length 512 for both μ and σ. The decoder g has a similar but reversed structure, where a vector of 512 is first passed through a linear layer. The ArcFace embedding is also concatenated on the channels after the same mapping. It goes through a convolution layer and then four functional blocks to restore the image size to 256×256. The first block is an Attention block, followed by three ResNet blocks. The numbers of channels for these blocks in the encoder are 64, 128, 128, and 256, and these numbers are reversed for the decoder blocks. However, since the encoder needs to accept both VIS and NIR images, each block within the encoder has two layers, while each block within the decoder has only one layer.

The loss functions for model training also originate from VAE but with some differences. Firstly, there is the reconstruction loss, which represents the error in restoring the image through the entire encoder and decoder chain. The reconstruction loss is calculated as the mean squared error between the original image and the restored image, but only the face region is considered in the calculation. It was done by using a facial landmark model to calculate the mask regions m_{nir} and m_{vis} from the face images x_{nir} and x_{vis}, respectively. Consequently, two reconstruction losses are obtained in Equations (10) and (11). The $\odot$ operator denotes the element-wise multiplication, which is used to mask the face region.

$$\mathcal{L}_{rec\_nir} = ||m_{nir} \odot (x_{nir} - x'_{nir})||^2 \tag{10}$$

$$\mathcal{L}_{rec\_vis} = ||m_{vis} \odot (x_{vis\_r} - x'_{vis\_r})||^2 \tag{11}$$

The second type of loss is the Kullback-Leibler (KL) loss, which in VAE is typically used to represent the distance between a distribution and the standard normal distribution $\mathcal{N}(0,1)$. Here, we only calculate this loss for VIS because VIS is the source image during inference. The KL loss term is written as $\mathcal{L}_{KL\_norm}$ in Equation (12).

$$\begin{aligned} \mathcal{L}_{KL\_norm} &= KL\left(\mathcal{N}(\mu_{vis}, \sigma^2_{vis}) || \mathcal{N}(0,1)\right) \\ &= -0.5\left(1 + \log\left(\sigma^2_{vis}\right) - \mu^2_{vis} - \sigma^2_{vis}\right) \end{aligned} \tag{12}$$

Additionally, we would like samples of the same identity to have their VIS and NIR latent vectors close to each other. To achieve this, we introduce an additional KL loss, as shown in Equation (13).

$$\begin{aligned} \mathcal{L}_{KL\_id} &= KL\left(\mathcal{N}\left(\mu_{vis}, \sigma^2_{vis}\right) || \mathcal{N}\left(\mu_{nir}, \sigma^2_{nir}\right)\right) \\ &= -0.5\left(1 - \log\left(\frac{\sigma^2_{nir}}{\sigma^2_{vis}}\right) - \frac{\sigma^2_{vis} + (\mu_{vis} - \mu_{nir})^2}{\sigma^2_{nir}}\right) \end{aligned} \tag{13}$$

The final loss function is a weighted sum of the reconstruction loss and the KL losses, as shown in Equation (14), where α and β are hyperparameters that control the weight of each KL loss.

$$\mathcal{L}_{total} = \mathcal{L}_{rec\_nir} + \mathcal{L}_{rec\_vis} + \alpha\mathcal{L}_{KL\_norm} + \beta\mathcal{L}_{KL\_id} \tag{14}$$

Inspired by CycleGAN [110], we have additionally introduced a cycle loss, as shown in Equation (15). However, since such implicit constraint losses often lead to training non-convergence, this loss was not introduced at the beginning of the training but was added after 2500 epochs of training with $\mathcal{L}_{total}$. The terms in loss $\mathcal{L}_{cycle}$ represent that a VIS image, after passing through the encoder f to obtain $(\mu_{vis}, \sigma_{vis})$, and then through the decoder g_{nir} to generate an NIR image x'_{nir}, is sent back to the encoder f to produce a latent vector pair $(\mu_{vis\_nir}, \sigma_{vis\_nir})$. We believe that μ_{vis} and μ_{vis_nir} should be as close as possible because they belong to different VIS and NIR domains of the same photo. The cycle loss is calculated as the mean squared error (MSE) between the two

latent vectors, μ_{vis} and μ_{vis_nir}. We did not consider the case of σ because σ may differ across different physical domains, hence we only used MSE loss instead of KL loss here.

$$\begin{aligned}(\mu_{vis}, \sigma_{vis}) &= f(x_{vis\_r}) \\ (\mu_{vis\_nir}, \sigma_{vis\_nir}) &= f(x'_{nir}) \\ \mathcal{L}_{cycle} &= ||\mu_{vis} - \mu_{vis\_nir}||^2\end{aligned} \tag{15}$$

Once the model is trained, we can use the target's video frames for inference. Each VIS image frame of the target's video will pass through the trained encoder f along with ArcFace f_{arc}, then the trained decoder g_{nir}. During inference, we do not sample from μ and σ but directly use μ as z input to g_{nir}. The process is depicted in Equation (16). The final output is a video consisting of a sequence of NIR image frames x'_{nir} generated from the sequence of VIS image frames x_{vis}.

$$\begin{aligned}(\mu_{vis}, \sigma_{vis}) &= f(x_{vis\_r}) \\ z_{id} &= f_{arc}(x_{vis}) \\ x'_{nir} &= g_{nir}(\mu_{vis}, z_{id})\end{aligned} \tag{16}$$

However, this method was found to severely overfit during our training. This is primarily attributed to the training dataset we could obtain, CASIA NIR-VIS, is too small, which contains only 725 subjects (including repeated identities). Generally, generative models require a larger amount of training data than classification models, and facial recognition models like ArcFace are even trained on datasets with millions of facial images. Therefore, overfitting on such a limited dataset is expected. To validate our conjecture, we included a small number of target samples in the training data and found that this could produce comparatively better results. This renders our method a few-shot learning approach, but we anticipate that with sufficient training data, the method could achieve zero-shot learning.

4.3 Results of the Red Bleed Attack Using Generative Samples

Instead of replaying a previously captured live NIR video, we synthesize an NIR video from an RGB video record of each volunteer, with the goal of spoofing the face authentication on the HP Windows Hello module. The experimental setup and evaluation results are detailed below.

Database and Preprocessing. The CASIA NIR-VIS 2.0 face database [51] comprises human face samples from 725 subjects. It is partitioned into four sessions, with each subject having between 1–22 VIS samples and 5–50 NIR samples. The original image resolution is 640×480. Due to image quality considerations, only the first three sessions are utilized in our generative model training. In addition, we collected nine NIR–VIS image pairs from each target subject for training and

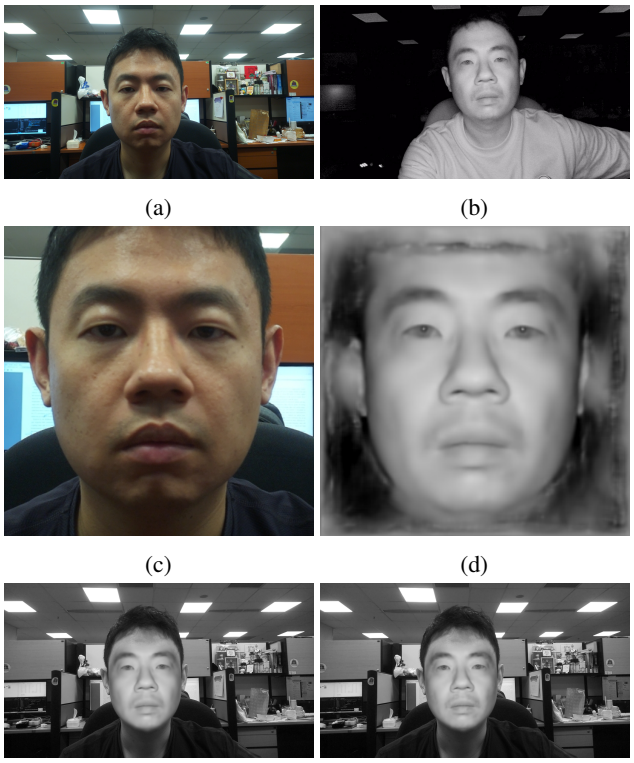


Figure 9: Image samples in the generative experiment. (a) VIS sample. (b) A ground-truth NIR sample with a similar pose. (c) The aligned VIS face using DeepFaceLab. (d) The synthesis result from the VAE with nine-shot learning. (e) The final generative image from nine-shot learning. (f) The final generative image from one-shot learning.

100 VIS samples for attack sample synthesis. The collected NIR images from OV9281 have a resolution of 1280×720, while the collected VIS images from the Raspberry Pi camera module 3 have a resolution of 2304×1296. Using DeepFaceLab [69], we apply an affine transformation to crop and align the full face region and generate the corresponding face mask. The aligned images measure 256×256 and serve as the input for both training and testing.

Experiment Setting. Similar to the setup described in Section 3.5, we displayed the generated video on our improvised NIR LCD to unlock the Windows Hello. The generative model was trained under two distinct protocols: (1) a nine-shot learning protocol, wherein nine NIR–VIS image pairs of the target subject (the volunteer registered in the system) were integrated into the CASIA NIR–VIS 2.0 face database, and (2) a one-shot learning protocol, wherein only one NIR–VIS image pair was added to the CASIA training set. The ArcFace feature embeddings are from ResNet-100 pretrained with MS1MV3 dataset. In our training phase, we set $\alpha = 0.01$ and $\beta = 0.01$. Both models were trained for 3,000 epochs at

Table 4: Attack success rates of Red Breed tested using samples generated from our trained generative model with nine-shot and one-shot learning.

Model	Successful rate
Nine-shot learning	97.73%
One-shot learning	61.25%

an initial learning rate of 0.0001 using the Adam optimizer with cosine annealing learning rate scheduler, after which we generated 100 NIR images from 100 VIS inputs.

Subsequently, the mask was applied to the generated NIR images to mitigate noise, and DeepFaceLab [69] was employed to transfer the generated NIR faces to the red channel of the original images. Default settings were retained, with the exception of an erode mask modifier of 15, a blur mask modifier of 72. Finally, the processed frames were concatenated into an FFV1 code video at a frame rate of 30 fps. Similarly to the setup described in Section 3.5, we displayed the generated video of volunteers on our improvised NIR LCD to unlock the HP Windows Hello module, then recorded the average success rate of 20 attempts of each volunteer in a day.

Evaluation Results. The used image samples and generation results are illustrated in Figure 9. The overall attack success rates for nine-shot and one-shot learnings are shown in Table 4. In addition, the attack video and the attack video record from the nine-shot learning are attached in the supplementary file [5]. In our nine-shot learning experiment, only the white American volunteer had failed to unlock the device in 10 out of 20 attempts. This may be attributed to the bias in the training data, as almost all subjects in the CASIA dataset are from Asia. As for the one-shot learning, though less effective, such an attack is shown to be feasible and pragmatic in the event that only limited sample pairs can be fetched. The one-shot generative sample in Figure 9f has a subtle difference compared with the nine-shot result shown in Figure 9e, and this is sufficient to affect the verification performance. The performance may be further improved if a better quality cross-spectral NIR-VIS face dataset with pixel correspondence is used; however, we cannot find such a dataset in the public domain at this juncture.

5 Suggested Countermeasures

Multimodal Authentication. When the enhanced face anti-spoofing feature is enabled in Windows Hello, our attack cannot bypass the facial authentication mechanism. This enhanced security measure requires a high-quality RGB image for validation, an element not provided by our attack. However, this enhanced feature increases authentication time and necessitates bright illumination, potentially rendering it impractical for certain access control scenes with poor lighting conditions, and the authentication failure is further aggregated by the increase in detection threshold upon every unsuccessful attempt by an authentic user. Likewise, other biometric systems can employ multimodal authentication for stringent security requirements, although this option generally involves additional sensors and extended authentication duration.

Active Projection or 3D Information Recovering. The Red Bleed attack is only feasible on single-eye cameras that lack robust 3D information capture. In contrast, sensors that project structured light, such as Apple FaceID, are capable of gathering 3D information for anti-spoofing, thereby thwarting the Red Bleed attack. Similarly, dual-camera systems that recover 3D geometry via multiview approaches can also defend against this attack. However, these countermeasures typically require additional hardware components and increase overall costs.

Further Investigation of Spatial-Temporal Features in NIR Domain. The Red Bleed attack is a newly introduced NIR replay attack in the biometric domain. In a manner analogous to the substantial research on RGB face anti-spoofing, researchers could further examine the unique characteristics of this technique. For instance, noticeable differences between the genuine sample and the attack sample, as shown in Figure 6, highlight potential artifacts that could be leveraged to detect and mitigate spoofing events.

6 Conclusion

We present the Red Bleed attack, a face presentation replay attack within the NIR domain, which exploits the vulnerability of Windows Hello's face authentication system under its default settings. Our experiments, using real captured face samples of 22 subjects of both genders, different races and age groups, demonstrate 100% attack success rates, thereby validating the effectiveness of this presentation attack. Furthermore, we introduce a VAE framework designed to convert visible image samples to NIR video samples, which can also successfully unlock the Windows Hello facial authentication mechanism with excellent success rate using few-shot learning and moderate success rate using one-shot learning. The attack gadget can be easily self-constructed from COTS materials, highlighting the compelling awareness of this vulnerability and necessity for both vendors and users to appropriately fortify existing plain-vanilla NIR biometric systems and exercise vigilance to mitigate the risk.

Acknowledgments

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) Imperial/NTU CYber Protection for HEalthcaRe (IN-CYPHER) programme, and the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2

Award No. MOET2EP50220-0003.

Ethics Considerations

Disclosure: This NIR face authentication vulnerability was discovered by surprise and sidetracked from our original research intent to explore polarization imaging for liveness detection against 3D face spoofing attacks. An immediate social impact arises for users relying on the default Windows Hello facial authentication: if their NIR or VIS images (or associated video samples) are compromised, the vulnerability detailed in this work can be exploited. Consequently, upon confirming the effectiveness of the attack on September 25, 2024, we immediately notified the security departments of Microsoft (MSRC), HP (HP PSRT), Dell (DELL PSIRT), and Lenovo (LENOVO PSIRT). We received their respective case numbers (MSRC 91421, HP-PSRT-IR 5182, DELL VRT-24281, LEN-174330) in prompt replies.

We subsequently collaborated with Microsoft MSRC to further examine the "enhanced anti-spoofing" feature, which, when enabled, can resist our Red Bleed attack presented in this paper. This information was also shared with the other three companies of which their products were used for our investigation. All parties inquired about our disclosure timeline, and upon being informed of our intention to submit this research to a leading security conference in January, we did not receive any objection about this submission.

We provided complete technical details in a report to each vendor and addressed their questions. After verifying our findings, HP, Dell, and Lenovo deferred to Microsoft MSRC to coordinate and implement a comprehensive solution. In the communication with MSRC on January 18, 2025, they confirmed our described behavior. At Microsoft's request, we shared our submitted conference paper. In their latest correspondence on January 20, 2025, Microsoft indicated that they would fully address this vulnerability in the June 2025 update. In accordance with responsible disclosure practices, we have agreed to withhold public release of this research until Microsoft has fully mitigated the issue. Microsoft managed to address this vulnerability earlier than expected with the CVE-2025-26644 published and the patch released in the KB5055523 security update for all in-service Windows versions on April 8, 2025.

Data Collection and Privacy: In the data collection process for this study, 22 volunteers participated. These volunteers provided explicit, written consent authorizing the academic use of their facial biometric data. All participants involved in conducting the evaluation of these vulnerabilities were fully briefed on the study's objectives and potential risks. For the privacy and ethical concerns, we deleted the registration of each individual volunteer immediately after the evaluation in his/her presence. The volunteers' biometric data will not appear in the public domain except for one volunteer who has signed the agreement explicitly. Moreover, we have signed a formal agreement stipulating that the dataset used will be used exclusively for academic purposes, in accordance with established ethical guidelines.

Negative Outcomes: Although Microsoft remediated this vulnerability in the Windows update on April 8, 2025, many deployed devices, particularly air-gapped devices, experience delayed patching. Since this vulnerability requires no internet connection to exploit, we strongly recommend all potentially affected users immediately upgrade to the latest system version.

There are also traditional biometrics relying on the NIR imaging, such as iris and palmprint [16, 46, 106]. In the absence of more robust anti-spoofing (liveness detection) measures, or only elementary spatial-temporal strategies are employed, these systems are similarly susceptible to the Red Bleed attack based on the same principal assumption of their security. The vulnerabilities exposed in this study may be inherently difficult to remediate in these existing biometric systems. Publication may render existing devices susceptible to exploits, with potential real-world consequences. We strongly urge affected vendors to issue public security advisories, implement proactive mitigation measures, and accelerate the phase-out of vulnerable devices.

Positive Outcomes: This work clearly identified the root cause of an intrinsic vulnerability of NIR-based biometric authentication that allows for possible fortification to be developed. The disclosure of this long-lasting security myth can raise awareness and increase the vigilance on the use of legacy plain vanilla NIR-based biometric authentication before malicious actors discover this attack method, which may result in greater harms if the finding is kept private. We anticipate that this work will inspire the development of more secure anti-spoofing methodologies, thereby enhancing the security posture of future biometric systems.

Open Science

Since the vulnerability CVE-2025-26644 has been made public, and the fix was conducted in the latest update, we publicly release all the code and hardware design details with documentation at Zenodo to the community for further research. Microsoft has independently reproduced the results and confirmed the effectiveness following the provided documents and video demos [5] of our initial successful attack setup.

The vulnerability has been patched in the latest Windows update. For attack evaluation purposes, we recommend using pre-April 8, 2025 system versions. The release of codes and hardware design details not only ensures transparency and reproducibility but also helps to promote the development of more secure and efficient biometric authentication by the research community.

References

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[2] Waqar Ali, Wenhong Tian, Salah Ud Din, Desire Iradukunda, and Abdullah Aman Khan. Classical and modern face recognition approaches: a complete review. *Multimedia tools and applications*, 80:4825–4880, 2021.

[3] Sizhe An, Hongyi Xu, Yichun Shi, Guoxian Song, Umit Y Ogras, and Linjie Luo. Panohead: Geometry-aware 3d full-head synthesis in 360deg. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20950–20959, 2023.

[4] Yousef Atoum, Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Face anti-spoofing using patch and depth-based cnns. In *2017 IEEE international joint conference on biometrics (IJCB)*, pages 319–328. IEEE, 2017.

[5] Red Bleed attack. Supplementary files in google drive, 2025. https://drive.google.com/drive/folders/1LDAiqCM8cCWiOFXTpTaiB5Ztcyyryu1k?usp=sharing.

[6] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 843–852, 2023.

[7] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1578–1587, 2022.

[8] Thomas Brox and Jitendra Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):500–513, 2010.

[9] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022.

[10] Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5799–5809, 2021.

[11] Haonan Chen, Guosheng Hu, Zhen Lei, Yaowu Chen, Neil M Robertson, and Stan Z Li. Attention-based two-stream convolutional networks for face spoofing detection. *IEEE Transactions on Information Forensics and Security*, 15:578–593, 2019.

[12] Li Chen, Mengyi Zhao, Yiheng Liu, Mingxu Ding, Yangyang Song, Shizun Wang, Xu Wang, Hao Yang, Jing Liu, Kang Du, et al. Photoverse: Tuning-free image customization with text-to-image diffusion models. *arXiv preprint arXiv:2309.05793*, 2023.

[13] Zhuowei Chen, Shancheng Fang, Wei Liu, Qian He, Mengqi Huang, Yongdong Zhang, and Zhendong Mao. Dreamidentity: Improved editability for efficient face-identity preserved image generation. *arXiv preprint arXiv:2307.00300*, 2023.

[14] J Choi, S Kim, Y Jeong, Y Gwon, and S Yoon. Conditioning method for denoising diffusion probabilistic models. *DOI: https://doi. org/10.1109/iccv48922*, 2021.

[15] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[16] John Daugman. How iris recognition works. In *The essential guide to image processing*, pages 715–739. Elsevier, 2009.

[17] Debayan Deb and Anil K Jain. Look locally infer globally: A generalizable face anti-spoofing approach. *IEEE Transactions on Information Forensics and Security*, 16:1143–1157, 2020.

[18] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[19] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[20] Zheng Ding, Xuaner Zhang, Zhihao Xia, Lars Jebe, Zhuowen Tu, and Xiuming Zhang. Diffusionrig: Learning personalized priors for facial appearance editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12736–12746, 2023.

[21] Chi Nhan Duong, Thanh-Dat Truong, Khoa Luu, Kha Gia Quach, Hung Bui, and Kaushik Roy. Vec2face: Unveil human faces from their blackbox features in face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6132–6141, 2020.

[22] edmundoptics. High contrast linear polarizing film (xp42), 2025.

[23] edmundoptics. Polymer polarizers and retarders, 2025.

[24] Florian M Farke, Leona Lassak, Jannis Pinter, and Markus Drmuth. Exploring user authentication with windows hello in a small business environment. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 523–540, 2022.

[25] Litong Feng, Lai-Man Po, Yuming Li, Xuyuan Xu, Fang Yuan, Terence Chun-Ho Cheung, and Kwok-Wai Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, 38:451–460, 2016.

[26] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022.

[27] Rinon Gal, Moab Arar, Yuval Atzmon, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. Encoder-based domain tuning for fast personalization of text-to-image models. *ACM Transactions on Graphics (TOG)*, 42(4):1–13, 2023.

[28] Javier Galbally and Sébastien Marcel. Face anti-spoofing based on general image quality assessment. In *2014 22nd international conference on pattern recognition*, pages 1173–1178. IEEE, 2014.

[29] Anjith George and Sébastien Marcel. Deep pixel-wise binary supervision for face presentation attack detection. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

[30] Anjith George and Sébastien Marcel. On the effectiveness of vision transformers for zero-shot face anti-spoofing. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021.

[31] Ying Guo, Xingxing Wei, Guoqiu Wang, and Bo Zhang. Meaningful adversarial stickers for face recognition in physical world. *arXiv e-prints*, pages arXiv–2104, 2021.

[32] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[33] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

[34] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[35] Xin Huang, Ruizhi Shao, Qi Zhang, Hongwen Zhang, Ying Feng, Yebin Liu, and Qing Wang. Humannorm: Learning normal diffusion model for high-quality and realistic 3d human generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4568–4577, 2024.

[36] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. *Pattern Recognition*, 115:107888, 2021.

[37] Minguk Kang, Joonghyuk Shin, and Jaesik Park. Studiogan: a taxonomy and benchmark of gans for image synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[38] Minguk Kang, Jun-Yan Zhu, Richard Zhang, Jaesik Park, Eli Shechtman, Sylvain Paris, and Taesung Park. Scaling up gans for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10124–10134, 2023.

[39] Manuel Kansy, Anton Raël, Graziana Mignone, Jacek Naruniec, Christopher Schroers, Markus Gross, and Romann M Weber. Controllable inversion of black-box face recognition models via diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3167–3177, 2023.

[40] Tero Karras. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948*, 2019.

[41] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances*

in neural information processing systems, 34:852–863, 2021.

[42] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[43] Taewook Kim, YongHyun Kim, Inhan Kim, and Daijin Kim. Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[44] Tobias Kirschstein, Simon Giebenhain, and Matthias Nießner. Diffusionavatars: Deferred diffusion for high-fidelity 3d head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5481–5492, 2024.

[45] Stepan Komkov and Aleksandr Petiushko. Advhat: Real-world adversarial attack on arcface face id system. In *2020 25th international conference on pattern recognition (ICPR)*, pages 819–826. IEEE, 2021.

[46] Adams Kong, David Zhang, and Mohamed Kamel. A survey of palmprint recognition. *pattern recognition*, 42(7):1408–1418, 2009.

[47] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2023.

[48] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(7):1794–1809, 2018.

[49] Lei Li, Zhaoqiang Xia, Abdenour Hadid, Xiaoyue Jiang, Haixi Zhang, and Xiaoyi Feng. Replayed video attack detection based on motion blur analysis. *IEEE Transactions on Information Forensics and Security*, 14(9):2246–2261, 2019.

[50] Lei Li, Zhaoqiang Xia, Xiaoyue Jiang, Yupeng Ma, Fabio Roli, and Xiaoyi Feng. 3d face mask presentation attack detection based on intrinsic image analysis. *Iet Biometrics*, 9(3):100–108, 2020.

[51] Stan Z Li, Dong Yi, Zhen Lei, and Shengcai Liao. The casia nir-vis 2.0 face database. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 348–353. IEEE, 2013.

[52] Xiaobai Li, Jukka Komulainen, Guoying Zhao, Pong-Chi Yuen, and Matti Pietikäinen. Generalized face anti-spoofing by detecting pulse from face videos. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4244–4249. IEEE, 2016.

[53] Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, and Ying Shan. Photomaker: Customizing realistic human photos via stacked id embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8640–8650, 2024.

[54] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 389–398, 2018.

[55] Yaojie Liu and Xiaoming Liu. Physics-guided spoof trace disentanglement for generic face anti-spoofing. *arXiv preprint arXiv:2012.05185*, 2020.

[56] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep tree learning for zero-shot face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4680–4689, 2019.

[57] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong. Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing. In *European Conference on Computer Vision*, pages 511–528. Springer, 2022.

[58] Oeslle Lucena, Amadeu Junior, Vitor Moia, Roberto Souza, Eduardo Valle, and Roberto Lotufo. Transfer learning using convolutional neural networks for face anti-spoofing. In *Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings 14*, pages 27–34. Springer, 2017.

[59] Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence*, 41(5):1188–1202, 2018.

[60] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

[61] Dann Mensah, Nam Hee Kim, Miika Aittala, Samuli Laine, and Jaakko Lehtinen. A hybrid generator architecture for controllable face synthesis. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–10, 2023.

[62] MicroSoft. Windows hello face authentication, 2021.

[63] Microsoft Support. Configure windows hello, n.d. Accessed: 2025-01-16.

[64] Usman Muhammad, Tuomas Holmberg, Wheidima Carneiro de Melo, and Abdenour Hadid. Face antispoofing via sample learning based recurrent neural network (rnn). In *The British Machine Vision Conference 2019 (BMVC) 9th-12th September 2019, Cardiff UK*. British Machine Vision Association Press, 2019.

[65] Xichen Pan, Li Dong, Shaohan Huang, Zhiliang Peng, Wenhu Chen, and Furu Wei. Kosmos-g: Generating images in context with multimodal large language models. *arXiv preprint arXiv:2310.02992*, 2023.

[66] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2face: A foundation model of human faces. *arXiv preprint arXiv:2403.11641*, 2024.

[67] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, and Stefanos Zafeiriou. Relightify: Relightable 3d faces from a single image via diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8806–8817, 2023.

[68] Xu Peng, Junwei Zhu, Boyuan Jiang, Ying Tai, Donghao Luo, Jiangning Zhang, Wei Lin, Taisong Jin, Chengjie Wang, and Rongrong Ji. Portraitbooth: A versatile portrait model for fast identity-preserved personalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27080–27090, 2024.

[69] Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Carl Shift Facenheim, Luis RP, Jian Jiang, Sheng Zhang, et al. Deepfacelab: Integrated, flexible and extensible face-swapping framework. *arXiv preprint arXiv:2005.05535*, 2020.

[70] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.

[71] Le Qin, Fei Peng, Min Long, Raghavendra Ramachandra, and Christoph Busch. Vulnerabilities of unattended face verification systems to facial components-based presentation attacks: An empirical study. *ACM Transactions on Privacy and Security*, 25(1):1–28, 2021.

[72] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[73] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

[74] Anton Razzhigaev, Klim Kireev, Edgar Kaziakhmedov, Nurislam Tursynbek, and Aleksandr Petiushko. Black-box face recovery from identity features. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 462–475. Springer, 2020.

[75] Yasar Abbas Ur Rehman, Lai-Man Po, Mengyang Liu, Zijie Zou, and Weifeng Ou. Perturbing convolutional feature maps with histogram of oriented gradients for face liveness detection. In *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019) Seville, Spain, May 13th-15th, 2019 Proceedings 12*, pages 3–13. Springer, 2020.

[76] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[77] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510, 2023.

[78] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, and Kfir Aberman. Hyperdreambooth: Hypernetworks for fast personalization of text-to-image models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6527–6536, 2024.

[79] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

[80] Martin Schadt. Liquid crystal materials and liquid crystal displays. *Annual review of materials science*, 27(1):305–379, 1997.

[81] Katja Schwarz, Axel Sauer, Michael Niemeyer, Yiyi Liao, and Andreas Geiger. Voxgraf: Fast 3d-aware image synthesis with sparse voxel grids. *Advances in Neural Information Processing Systems*, 35:33999–34011, 2022.

[82] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 acm sigsac conference on computer and communications security*, pages 1528–1540, 2016.

[83] Meng Shen, Zelin Liao, Liehuang Zhu, Ke Xu, and Xiaojiang Du. Vla: A practical visible light-based attack on face recognition systems in physical world. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–19, 2019.

[84] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[85] Wenyun Sun, Yu Song, Changsheng Chen, Jiwu Huang, and Alex C Kot. Face spoofing detection based on local ternary label supervision in fully convolutional networks. *IEEE Transactions on Information Forensics and Security*, 15:3181–3196, 2020.

[86] C Szegedy. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[87] Thanh-Dat Truong, Chi Nhan Duong, Ngan Le, Marios Savvides, and Khoa Luu. Vec2face-v2: Unveil human faces from their blackbox features via attention-based network in face recognition. *arXiv preprint arXiv:2209.04920*, 2022.

[88] Omer Tsarfati. Bypassing windows hello without masks or plastic surgery, 2023.

[89] Dani Valevski, Danny Lumen, Yossi Matias, and Yaniv Leviathan. Face0: Instantaneously conditioning a text-to-image model on a face. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023.

[90] Qinghe Wang, Xu Jia, Xiaomin Li, Taiqing Li, Liqian Ma, Yunzhi Zhuge, and Huchuan Lu. Stableidentity: Inserting anybody into anywhere at first sight. *arXiv preprint arXiv:2401.15975*, 2024.

[91] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, Anthony Chen, Huaxia Li, Xu Tang, and Yao Hu. Instantid: Zero-shot identity-preserving generation in seconds. *arXiv preprint arXiv:2401.07519*, 2024.

[92] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[93] Zhenqi Xu, Shan Li, and Weihong Deng. Learning temporal features using lstm-cnn architecture for face anti-spoofing. In *2015 3rd IAPR asian conference on pattern recognition (ACPR)*, pages 141–145. IEEE, 2015.

[94] Yueli Yan and Zhice Yang. Spoofing real-world face authentication systems through optical synthesis. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 882–898. IEEE, 2023.

[95] Yuxuan Yan, Chi Zhang, Rui Wang, Yichao Zhou, Gege Zhang, Pei Cheng, Gang Yu, and Bin Fu. Facestudio: Put your face everywhere in seconds. *arXiv preprint arXiv:2312.02663*, 2023.

[96] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.

[97] Ziqi Yang, Jiyi Zhang, Ee-Chien Chang, and Zhenkai Liang. Neural network inversion in adversarial setting via background knowledge alignment. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 225–240, 2019.

[98] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

[99] Bangjie Yin, Wenxuan Wang, Taiping Yao, Junfeng Guo, Zelun Kong, Shouhong Ding, Jilin Li, and Cong Liu. Adv-makeup: A new imperceptible and transferable attack on face recognition. *arXiv preprint arXiv:2105.03162*, 2021.

[100] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Face anti-spoofing with human material perception. In *Computer Vision–ECCV 2020:*

16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, pages 557–575. Springer, 2020.

[101] Zitong Yu, Xiaobai Li, Pichao Wang, and Guoying Zhao. Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection. *IEEE Signal Processing Letters*, 28:1290–1294, 2021.

[102] Zitong Yu, Yunxiao Qin, Xiaobai Li, Chenxu Zhao, Zhen Lei, and Guoying Zhao. Deep learning for face anti-spoofing: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(5):5609–5631, 2022.

[103] Ge Yuan, Xiaodong Cun, Yong Zhang, Maomao Li, Chenyang Qi, Xintao Wang, Ying Shan, and Huicheng Zheng. Inserting anybody in diffusion models via celeb basis. *arXiv preprint arXiv:2306.00926*, 2023.

[104] Longwen Zhang, Qiwei Qiu, Hongyang Lin, Qixuan Zhang, Cheng Shi, Wei Yang, Ye Shi, Sibei Yang, Lan Xu, and Jingyi Yu. Dreamface: Progressive generation of animatable 3d faces under text guidance. *arXiv preprint arXiv:2304.03117*, 2023.

[105] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

[106] Dexing Zhong, Xuefeng Du, and Kuncai Zhong. Decade progress of palmprint recognition: A brief survey. *Neurocomputing*, 328:16–28, 2019.

[107] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, and Lizhuang Ma. Test-time domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 175–187, 2024.

[108] Yufan Zhou, Ruiyi Zhang, Tong Sun, and Jinhui Xu. Enhancing detail preservation for customized text-to-image generation: A regularization-free approach. *arXiv preprint arXiv:2305.13579*, 2023.

[109] Zhe Zhou, Di Tang, Xiaofeng Wang, Weili Han, Xiangyu Liu, and Kehuan Zhang. Invisible mask: Practical attacks on face recognition with infrared. *arXiv preprint arXiv:1803.04683*, 2018.

[110] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.