BioGaze: a Framework for Evaluating the Photographic Requirements of the ISO/IEC 39794-5 Standard

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Elatfi, O., Di Domenico, N., Borghi, G., Franco, A., Maltoni, D. (2025). BioGaze: a Framework for Evaluating the Photographic Requirements of the ISO/IEC 39794-5 Standard. 345 E 47TH ST, NEW YORK, NY 10017 USA : Institute of Electrical and Electronics Engineers Inc. [10.1109/fg61629.2025.11099423].

Availability:
This version is available at: https://hdl.handle.net/11585/1037551 since: 2026-01-16

Published:
DOI: http://doi.org/10.1109/fg61629.2025.11099423

BioGaze: a framework for evaluating the photographic requirements of the ISO/IEC 39794-5 standard

Osama Elatfi[1], Nicolò Di Domenico[1], Guido Borghi[2], Annalisa Franco[1], Davide Maltoni[1]

[1] Department of Computer Science and Engineering, University of Bologna, Italy

[2] Department of Education and Humanities, University of Modena and Reggio Emilia, Italy

***Abstract*— Facial recognition is a key biometric technology, especially for using electronic documents in real-world applications. The accuracy of this recognition technology strictly depends on the image quality, *i.e.* the face appearance in the image included in the document. Then, adherence to ISO/ICAO standards, which contain guidelines to standardize the image quality in official documents, is of paramount importance. However, ensuring compliance is challenging due to high subject variability. Furthermore, controls are often executed manually, making them subjective and time-consuming. Therefore, in this work, we introduce BioGaze, an automated framework for ISO/ICAO compliance verification that combines classical computer vision and deep learning algorithms to perform the checks contained in the latest standard version. The framework is tested on a synthetic dataset, achieving state-of-the-art performance across multiple ISO/ICAO requirements, surpassing public algorithms and commercial SDKs. BioGaze is publicly available to advance automated compliance verification and support standardization efforts[1].**

I. Introduction

Facial recognition [43] has emerged as a key biometric technique for identifying and authenticating an individual, mostly thanks to its non-invasive nature and rapid acquisition times [9]. These two characteristics combined led to its placement at the forefront of security and identification applications, particularly in the context of electronic Machine-Readable Travel Documents (eMRTDs) such as electronic passports, where accurate facial recognition is essential to ensure border security [5].

However, the effectiveness of such biometric systems is heavily dependent on the standardization of image quality and format [14], as these two factors deeply influence the accuracy of facial recognition algorithms. To address these challenges, international organizations such as the International Civil Aviation Organization (ICAO) and the International Organization for Standardization (ISO) have established guidelines [21], [22], here collectively referred to as ISO/ICAO standards, to ensure interoperability and reliability across biometric systems in all jurisdictions.

From a historical point of view, ICAO was the first, in the 1980s, to attempt standardizing automatic biometric identification [39]. Subsequently, in line with these ICAO directives, ISO/IEC proposed the ISO/IEC 19794-5 [21] standard, later revised into ISO/IEC 39794-5 [22], defining the rules for encoding, recording, and transmitting facial images in electronic passports. Together, these standards aim to promote consistency and interoperability across different biometric systems, as well as outlining recommendations to ensure the quality of captured facial images. In particular, ISO/ICAO standards include a set of general guidelines, including specifications for head positioning, image dimensions, minimum inter-eye distance, background color and uniformity, facial lighting and tone, head and shoulder alignment, expression neutrality, eye visibility, and the presence of accessories such as glasses or headgear.

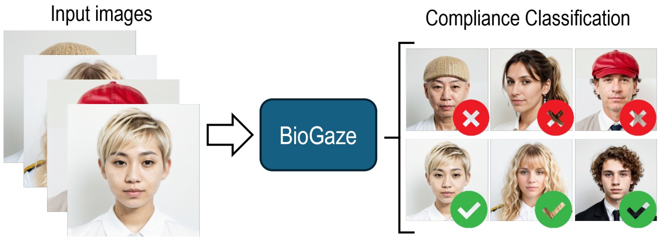


Fig. 1: Overview of the addressed task. We propose a framework, BioGaze, able to accurately classify if input images are compliant w.r.t. the ISO/IEC 39794-5 standard [22] for the face image quality.

Although these requirements are essential to maximize the accuracy of facial recognition algorithms, verifying the compliance of facial images with these guidelines is challenging [14]. This difficulty stems mainly from the high variability between subjects due to ethnic, cultural, and religious differences, making it difficult to create methods that can generalize well in all subjects. For instance, detecting the presence of shadows or occlusions on the face can be challenging due to skin color or the presence of the hijab or other cultural or religious ornaments. Another challenge is posed by the quality of the sensor used to capture the image in the photo booth, as lower-quality sensors can produce images with artifacts such as blurriness, poor contrast, or uneven lighting, further complicating the verification process.

Traditionally, the compliance verification process has been done manually by human examiners, such as border guards or officials in charge of issuing passports, and only partially through automated systems. However, manual verification presents some significant drawbacks: firstly, human assessment is inherently subjective, which can lead to variability

This project received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No.101121280. This text reflects only the author's views, and the commission is not liable for any use that may be made of the information contained therein.

[1] https://github.com/MI-BioLab/BioGaze

and possible inconsistencies in compliance evaluations. Furthermore, issues such as fatigue, distractions, and the repetitive nature of the task can compromise both the accuracy and reliability of the officer involved. Finally, manual inspections take a considerable amount of time, making them unsuitable for high-volume situations, such as busy border checkpoints or passport issuance centers.

In contrast, an automated solution would bring several advantages: by automating the compliance verification process, it would be possible to process more images while providing consistent, objective, and explainable metrics to support decision making, while also standardizing evaluation across countries. In addition, automated systems could also be installed in kiosks to provide immediate feedback during the image acquisition phase, assisting users to capture compliant images and reducing the likelihood of rejection and the need for retakes. Therefore, in recent years, there has been a shift towards employing specialized software systems, such as commercial SDKs, to at least partially automate this process.

However, despite the importance of automating the ISO/ICAO compliance verification workflow, the development and testing of automated systems has not been fully investigated in the literature. This can be mostly attributed to a substantial lack of both commercial and open-source software, which generally exhibit inconsistent performance across various checks, while others are not implemented at all [6] since ISO/ICAO are expressed in terms of guidelines and usually are not specified in detail. Moreover, the limited availability of facial images for training such systems, mostly due to privacy concerns [7], [6], poses challenges in the development of novel tools.

Therefore, in this work, we attempt to address the former issue by presenting and publicly releasing a novel ISO/ICAO compliance verification tool, namely BioGaze. This framework implements the checks proposed in the ISO/ICAO standards by employing a combination of classical computer vision and deep learning-based algorithms, in order to achieve high performance to classify compliant images (see Fig. 1). Using TONO [6], a recent synthetic dataset introduced in the literature, we demonstrate that BioGaze is able to achieve state-of-the-art results on a significant number of requirements deduced from the last ISO/ICAO standard, outperforming both publicly available algorithms, as well as several commercial off-the-shelf SDKs.

II. Related Works

A. Face quality and ISO/ICAO Compliance Verification

The impact of face image quality on recognition performance has driven the need to standardize this concept over time, to identify the key factors influencing image quality and codify them, including the development of algorithms to evaluate these aspects systematically. Ensuring high-quality images supports robust recognition performance, even in the presence of the aging effects typical of this domain, given the long-term validity period of identity documents.

The first guidelines in this sense were provided by ICAO Doc 9303 [19], which specifies the functional requirements for MRTDs and defines their key properties. The portrait printed on ICAO-compliant MRTDs is a crucial element, serving as a primary link between the document and its holder. High-quality standardized portraits facilitate identity verification by issuing agencies and enable both manual and automated inspection at borders. Since the introduction of digitally stored images in 2005, Automated Border Control (ABC) systems have been implemented to compare the printed and electronically stored images with live-captured images during border crossings, streamlining the verification process. The ICAO guidelines have been used as the reference for the development of ISO/IEC 19794-5 [21] standard, and its later evolution ISO/IEC 39794-5 [22].

Given the importance and impact of face image quality in more general applications [10], an additional standardization effort has been dedicated to addressing face image quality from a broader perspective. This effort aims to bridge the gap that has existed for years compared to other biometric traits (*e.g.*, fingerprints). In particular, the ISO/IEC 29794-5 [23] standard, currently under development, focuses on a general framework for assessing face image quality, suitable for a wide range of use cases where face images can be acquired in unsupervised conditions, resulting in significant variability in pose, lighting, and other factors.

B. ISO/ICAO Compliance Verification Tools

Following the ISO/ICAO guidelines, various tools for compliance verification have been developed over time. One of the first research tool in this domain was introduced in [14], where a comprehensive list of requirements was systematically defined for the first time. The study also proposed algorithmic solutions capable of verifying each individual requirement. This work served as a starting point for establishing an evaluation benchmark on the FVC-onGoing platform. Following the definition of the guidelines for ISO/ICAO compliance checks, some commercial solutions have been proposed. These include either tools specifically designed for ICAO verification or facial recognition SDKs that have extended their functionalities to include ICAO compliance checks. These solutions are complemented by a limited number of research contributions, which are open source in some cases. Before reviewing the main existing tools, it is important to highlight a key consideration. Although all tools are based on the same ISO/ICAO guidelines, each SDK implements a specific set of checks, which sometimes do not cover all the requirements. This variability makes it challenging to compare different systems, as establishing a precise mapping is often complicated by the differing interpretations of the guidelines adopted by each solution.

Regarding the existing commercial tools, an interesting SDK is developed by Correlance[2], which focuses on verifying compliance with the first version of the standard. Another commercial product is a general-purpose software from Innovatrics[3], which offers a more general approach that

[2]https://www.correlance.com/cms/en/ccEngineICAO
[3]https://www.innovatrics.com/digital-onboarding-toolkit/face-matching

includes checks for passive liveness and match quality. By returning immediate feedback to the user, this tool guides them through the acquisition process to meet the requirements. Similarly, the SDK provided by NEUROtechnology[4] focuses on basic recommendations for facial image quality, including minimum eye distance and checks for near-frontal faces. As the above-mentioned tools are commercial products, often part of more comprehensive SDKs for facial analysis and recognition, neither the source code nor their internal algorithms are disclosed. Therefore, their utility for research purposes is limited.

In addition to commercial solutions, research initiatives have proposed datasets and models to tackle ISO/ICAO compliance verification. A notable example is the BioLab-ICAO [30] framework, further refined in [14], which provides a comprehensive ground-truth dataset, a standardized testing protocol, and baseline algorithms for compliance verification. However, researchers are granted access to only a limited subset of images for training purposes, as the majority remain inaccessible and are used primarily as a benchmark on the FVC-onGoing[5] platform. Moreover, the same authors also present in [14] the BioLab-ICAO-Check tool, which employs a variety of traditional and specific computer vision algorithms, such as the Prewitt operator [12] to detect pixelation and TSI [15] to evaluate the blurriness of an image. Some other research contributions have proposed solutions for a specific subset of ICAO checks: head coverings, also related to religious circumstances [17], pixelation, hair across eyes, veil over face and mouth opened [35], eyes analysis (eyes closed, red eyes and looking away) [3] head pose [2], mouth closedness, eyes openness, and no veil-over-face [33]. Other tools have been designed to provide a wider coverage of the ISO/ICAO requirements. The authors of [38] introduce a novel integrated system for ISO/ICAO face image compliance, leveraging a method inspired by human brain functionality. The key contribution lies in using a unified approach for assessing multiple compliance requirements (nine of them are considered), unlike other existing methods that handle each requirement separately.

A more comprehensive solution is proposed in [13] where the authors propose ICAONet, a deep learning-based multi-task network for the automatic evaluation of photographic requirements of the ISO/IEC 19794-5 [21] standard. The network is composed of three main components. The first, the encoder part of the autoencoder presented in [16], is responsible for computing the embedding representation of the given face image. This embedding is then fed into two distinct networks, *i.e.* an unsupervised decoder branch that is trained to reconstruct the original image, and a supervised classification branch that predicts the scores for each requirement to verify. The proposed model has achieved state-of-the-art results, outperforming both academic and private algorithms, and delivering the best Equal Error Rate for several photographic requirements defined in the standard.

[4] https://www.neurotechnology.com/face-verification-technical.html
[5] https://biolab.csr.unibo.it/FvcOnGoing

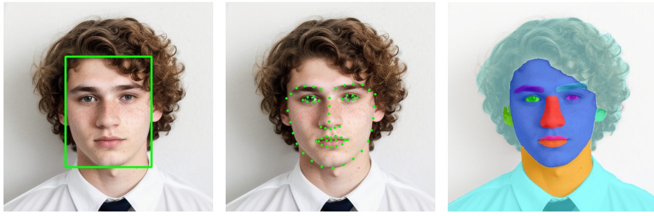


Fig. 2: The three fundamental steps on which is based the BioGaze framework, *i.e.* face detection, facial landmark estimation and face segmentation (see Sect. III-A).

III. BioGaze Framework

To address the task of automated ISO/ICAO compliance verification, we propose BioGaze, a tool that encompasses a mix of classical computer vision and deep learning-based algorithms, leveraging the strengths of both approaches. More specifically, the use of hand-crafted features provides explainability, making them ideal for certain tasks, while the robustness and generalization capabilities of state-of-the-art deep neural networks enhance performance in more complex scenarios. The framework is designed to take RGB images of variable size as input and return a score as output. The score is in the range $[0, 1]$ for each individual check to provide explanation and support to the decision-making process.

A variety of algorithms is implemented for each specific check deduced from the ISO standard. In the following, for each check, we provide a detailed analysis and motivations.

A. Detection of Face and Landmarks, Face Segmentation

There are three fundamental and preliminary operations performed for each input image. Face detection is used to understand if the input image really contains a face, and also to provide the face location information to subsequent operations. Facial landmark estimation provides the locations of important points on the face representing, for instance, the location of the pupils or the lips. These points are widely used in further checks to control the state of specific element, *e.g.* the degree of eye or mouth openness. Finally, face segmentation is necessary to identify the main image and facial components (e.g. background, hair, face, eyes, etc.). Summarizing, all these three tasks are important for subsequent checks and are tackled through deep learning networks. An example of the output of these initial steps is provided in Figure 2.

For the face detection task, we choose YOLOv8 [37] as our face detector, thanks to its high detection speed and accuracy. Considering that an ISO/ICAO compliance verification tool is likely to operate in a controlled setting, *i.e.* not with "in the wild" images, we believe the face detection task is not critical, and then face detection models can be changed without significant consequences.

For the facial landmark estimation task, we rely on the well-known Dlib [26], [25] library. In this case, we decide to detect landmarks with one of the most used library in the biometrics fields, improving then the interoperability of BioGaze with other future systems.

Finally, we employ the BiSeNet [41] segmentation network, adapted to tackle the face segmentation task by training on the CelebAMask-HQ [27] dataset, which provides high-quality annotations for the different facial regions, as well as covering headwear, eyeglasses, necklaces, and clothing.

B. Frontal pose of head and shoulders

To ensure the subject's face is fully visible, the standard defines the face and the shoulders must be in a frontal positon – specifically, yaw, roll and pitch angles must be in the range $[-5°, +5°]$ – with respect to the acquisition device. We address these two controls through two different algorithms, whose scores are then averaged together.

More specifically, the idea behind the first one is that a subject with frontal pose should have both shoulders equally visible, without any large disproportion between the two. To implement it, we leverage the segmentation map produced during the face parsing step; next, we count the number of shoulder pixels that are on the left and on the right half of the image, and the minimum of the two is divided by the maximum of the two. The second check leverages the approach available in MediaPipe [24], [29] to estimate the face's rotation angles. Using its ability to extract three-dimensional facial landmarks, we map these three-dimensional points to their 2D projections on the image plane. The face pose is then computed by solving a pose estimation problem [31] using the Levenberg-Marquardt [28], [32] optimization algorithm. Next, the resulting rotation matrix is decomposed using the Rodrigues formula [8] to extract the individual pitch (θ), yaw (ψ), and roll (ϕ) components. Each angle is then normalized so that the minimum and maximum acceptable angles correspond to -1 and 1. For instance, the angle θ is normalized θ_n as:

$$\theta_n = \frac{2\theta - \theta_{max} - \theta_{min}}{\theta_{max} - \theta_{min}} \tag{1}$$

Similarly, ψ_n and ϕ_n are normalized using the same formula. Finally, the score for this check S_{FP} is calculated as follows:

$$S_{\text{FP}} = 1 - 0.5 \cdot \sqrt{\theta_n^2 + \psi_n^2 + \phi_n^2} \tag{2}$$

This way, we obtain a score which is then normalized between 0 and 1, where high values represent lower angles and therefore better compliance.

C. Front-facing gaze

In ISO/ICAO-compliant images, subjects are required to face the camera directly, maintaining an upright head position. Specifically, the alignment of the eyes with the camera must be ensured.

We verify this requirement by relying on a ResNet-50 [18] trained on the ETH-XGaze [42] dataset which, given the input image, outputs the horizontal component of the gaze θ. We compute the final score S_{FG} as follows:

$$S_{\text{FG}} = 1 - \max\left(0, \min\left(\frac{\theta - \theta_l}{\theta_r - \theta_l}, 1\right)\right) \tag{3}$$

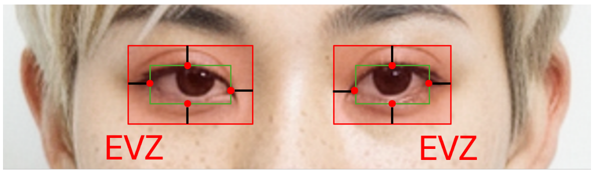


Fig. 3: Visualization of the Eye Visibility Zone (EVZ) [39], described as the rectangular zone encircling the eye, with a minimum distance of 5% of the inter-eye distance (IED) from visible sections of the eyeball (see Sect. III-D).

The value is normalized between 0 and 1 using two predefined values as left and right boundaries, respectively $\theta_l = -0.15$ and $\theta_r = 0.08$.

D. Eye Visibility

The standard defines that the eyes, one of the most relevant elements for face recognition, must be fully visible and naturally opened. In particular, an Eye Visibility Zone (EVZ) [39], *i.e.* a small rectangular area surrounding the eyes (see Fig. 3), is defined, and all elements (*e.g.* iris, pupil, eye socket) inside must not be occluded.

This requirement is then declined in three different checks.

1) Eyes open: As mentioned before, both eyes must be naturally opened. Then, an algorithm to determine the level of closure of eyes is needed. Following the approach depicted in [40], this check is implemented by computing the ratio between the aperture of the eyes and the inter-eye distance. Indeed, by leveraging the facial landmarks detected in the initial steps of the pipeline, we determine the maximum aperture for each eye, namely A_l and A_r, and select the smaller of the two values. This value is then divided by the inter-eye distance IED, which serves as a normalization factor to account for variations in the face size of the subject. The final score S_{EO} is then computed as depicted as follows:

$$S_{\text{EO}} = \sigma\left(\frac{\min(A_l, A_r)}{\text{IED}} - t\right) \tag{4}$$

where an empirical threshold $t = 0.09$ is subtracted from the normalized value and a sigmoid function σ is finally applied to map the result into a continuous range $[0, 1]$.

2) No unnatural makeup: the presence of makeup, especially if it is heavy and uses dark colors, can compromise the visibility of certain eye details and, in general, affect the performance of facial recognition technology. Therefore, it is necessary to determine whether the visibility of the EVZ is acceptable or not. This check is implemented by fusing the scores returned from two different algorithms. Initially, we calculate the EVZ, and we adapt this definition to our use case by adding a small extra margin to better cover the area around the eyes. The first algorithm starts by converting each EVZ into the HSV color space, and computing the normalized histogram on the hue channel. Since makeup typically occupies a hue range distinct from natural skin tones, we calculate the proportion of pixels with unnatural hues for each eye, denoted as H_l and H_r. Therefore, the score $S_{\text{MU-1}}$ of this first check is implemented as:

$$S_{\text{MU-1}} = 1 - (0.5 \cdot H_l + 0.5 \cdot H_r) \tag{5}$$

The second algorithm evaluates the discrepancy between the skin tone and the periocular region, based on the observation that makeup, especially in the area surrounding the eyes, often differs significantly from natural skin tones (in the case the makeup color is similar to the skin, this is accepted by the standard). This check is implemented by calculating the mean hue of the skin in the cheek area and comparing it to the mean hues of the two EVZs. To account for the circular nature of hue values in the HSV color space, the distance H_D between two hues H_a and H_b in the range [0, 1] is computed as:

$$H_D(H_a, H_b) = \min(|H_a - H_b|, 1 - |H_a - H_b|) \tag{6}$$

Using this metric, the algorithm computes the score as follows:

$$S_{\text{MU-2}} = 1 - (0.5 \cdot H_D(H_c, H_l) + 0.5 \cdot H_D(H_c, H_r)) \tag{7}$$

where H_c is the mean hue of the cheek area, and H_l and H_r are the mean hues of the left and right EVZs, respectively. Finally, the overall score S_{MU} is determined by averaging the outputs $S_{\text{MU-1}}$ and $S_{\text{MU-2}}$ of the two algorithms.

3) No sunglasses: The presence of sunglasses with dark lenses strongly hinders the visibility of the EVZ. Then, it is important to detect the presence of sunglasses.

This check focuses on analyzing the pixels associated with eyewear, as identified by the face segmentation map. Since sunglasses can exhibit a wide range of properties that affect brightness, *e.g.* opacity, variable colors, and reflections, we propose an approach that leverages the V channel of the HSV color space, which captures the brightness of the image.

First, the image is converted to the HSV color space, and then its normalized histogram on the V channel is computed. We observe that the presence of sunglasses is associated with higher values within the range $[20, 50]$. Then, the score S_{SG} for this check is computed as:

$$S_{\text{SG}} = 1 - \sum_{i=20}^{50} h_V(i) \tag{8}$$

where $h_V(i)$ represents the normalized histogram value at intensity i in the V channel.

E. Neutral expression

According to ISO/ICAO standards, subjects must maintain a neutral facial expression without smiling, even with a closed mouth. Additionally, the guidelines dictate that eyebrows should remain in their natural position, with squinting or frowning not allowed.

We verify this requirement by employing an ensemble of algorithms, both deep learning-based and landmark-based. The first one relies on the Residual Masking Network [36], a deep learning-based model for facial expression recognition that outputs the probability $S_{\text{NE-1}} = p_n$ that the face exhibits a neutral expression. The second approach, implemented following [22], calculates the score in the following manner:

$$S_{\text{NE-2}} = 1 - \min\left(1, \frac{2 \cdot d_{ul}}{t_l}\right) \tag{9}$$

where d_{ul} is the vertical distance between the midpoints of the upper and lower lips, and t_l is the thickness of the lower lip, that acts as a normalization value.

As we empirically observe that this method struggles to recognize subtle smiles, we implement a third method attempting to address this limitation. By analyzing facial landmarks, we count the number of pixels belonging to the lips (excluding teeth) n_L and divide this count by the total pixel count for the entire mouth region (including lips and teeth) n_M, thus obtaining $S_{\text{NE-3}} = n_L/n_M$. The output of these techniques is combined using a weighted mean to produce the final score as follows.

$$S_{\text{NE}} = \alpha \cdot S_{\text{NE-1}} + \beta \cdot S_{\text{NE-2}} + \gamma \cdot S_{\text{NE-3}} \tag{10}$$

The three weights $\alpha, \beta, \gamma = [0.2, 0.3, 0.5]$ are assigned a higher weight to checks that, on their own, yield better performance in terms of equal error rate.

F. Uniform face lighting

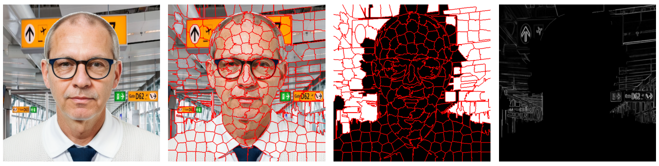
Following the standard, the light must be symmetrically distributed over the subject's face, creating an uniform illumination. Moreover, no visible reflections – for instance the ones produced by the flash light – should be observable in the skin.

We verify this requirement employing two distinct approaches. The first one is implemented following the guidelines reported in [22]. More specifically, the standard dictates that face illumination is deemed uniform if, for each channel, the minimum average intensity among four regions, *i.e.* forehead, left and right cheeks, and chin, is at least 50% of the maximum average intensity. Therefore, we calculate the numeric score $S_{\text{UL-1}}$ for this check by first computing a weighted mean:

$$\begin{aligned} S_G &= \alpha \cdot \frac{R_m}{R_M} + \beta \cdot \frac{G_m}{G_M} + \gamma \cdot \frac{B_m}{B_M} \\ S_{\text{UL-1}} &= \sigma(S_G - 0.5) \end{aligned} \tag{11}$$

where R_m, R_M, G_m, G_M, B_m, B_M are the minimum and maximum mean intensities for the three RGB color channels, respectively, multiplied with the coefficients to compute relative luminance from sRGB color space with values $\alpha, \beta, \gamma = [0.2126, 0.7152, 0.0722]$.

To obtain a check that is more robust to the presence of hair and occlusions in the four above-mentioned regions, we implement an additional method that computes the percentage of facial pixels in the grayscale image that are neither too bright (above 230) nor too dark (below 70). The resulting percentage, $S_{\text{UL-2}}$, is averaged with $S_{\text{UL-1}}$ to obtain the final score S_{UL} on which is applied a sigmoid function.



(a) Original (b) Superpixels (c) Mask (d) Mask. Grad.

Fig. 4: Visual sample of the different steps required to compute the score for the "Uniform background" check (see III-G). The input image (a) is segmented in superpixels as (b). Then, only the superpixels entirely contained in the background are kept, obtaining the mask (c). Finally, this mask is imposed on the image's gradient magnitudes (d).

G. Uniform background

According to ISO/ICAO standards, the background should be a plain, light-colored surface with a matte finish. It must be free from visible textures, such as spots, lines, or curves. Furthermore, the background lighting should be uniform, although gradual transitions from light to dark are permissible, provided they occur in only one direction.

Inspired by the methodology described in [34], this check consists of analyzing the mean gradient magnitude of the background. Specifically, a Sobel kernel is applied to compute the image gradient, from which its magnitude is extracted at each pixel. Firstly, the face parser segments the image, isolating the subject from the background. To further refine the background selection, the image is converted into the Lab color space, which separates lightness from color components. The L channel is subdivided with the SLIC [1] algorithm into 300 superpixels, which group adjacent pixels with similar characteristics into coherent regions. This segmentation helps reduce the amount of misclassified pixels by the face parser, which may negatively affect the result. Then, only those superpixels that are entirely enclosed within the background are retained, while any superpixels overlapping with the foreground are discarded. The final mask, composed of the remaining background superpixels, is applied to the gradient magnitude image, and the magnitudes of the remaining pixels are averaged. All these steps are visually summarized in Figure 4. Finally, the score for this check is obtained by subtracting a predefined threshold from the mean gradient magnitude and applying the sigmoid function.

H. Head without covering

The ISO/IEC 14496-2 [20] standard defines a series of facial landmarks, four of which can be used to define an ellipse. According to ISO/ICAO guidelines, the facial region enclosed in this ellipse must be fully visible and unoccluded by head coverings. To verify this requirement, we inspect the segmentation mask returned by the face parser in the initial steps of the pipeline. In particular, given n_h as the number of pixels classified as belonging to the "hat" class and n_f as the number of pixels classified as not belonging to the "background" class, the score of this check S_{HC} is computed as shown as follows:

$$S_{\text{HC}} = 1 - \frac{n_h}{n_f} \tag{12}$$

Therefore, we can obtain a score between 0 and 1, where 1 corresponds to the image containing zero pixels classified as "hat", *i.e.* the subject not wearing a hat.

I. Photographic Requirements

ISO/ICAO guidelines define rules and constraints to ensure that the subject's mugshot picture is captured in an optimal setting: specifically, images must be acquired in a manner that ensures an accurate depiction of colors, as well as being in focus. To ensure compliance with these guidelines, we perform several checks primarily focusing on color analysis, along with an additional check to determine whether the subject is in focus. More specifically, we analyze the image's colors by verifying exposure, saturation, and presence of both posterization and pixelation artifacts.

1) Correct exposure: we implement this check by examining the normalized grayscale histogram of the face region. Specifically, the analysis focuses on the intensity distribution of grayscale pixels within the face, with particular attention given to the darkest (0-170) and brightest (220-256) regions. If the average or maximum intensities in these areas exceed certain predetermined thresholds, the image is marked as either underexposed or overexposed.

2) Correct saturation: this check is implemented by first converting the image into the HSV color space and then analyzing the saturation values of the face region, as determined by the face parser. Based on empirical observations of the saturation histograms, we establish thresholds to classify pixels with too high (values above 200) and too low (values below 40) saturation. Therefore, we compute the score of this check as the percentage of pixels whose saturation falls in the correct range.

3) No posterization: this check aims to detect the presence of posterization in facial images, a visual artifact characterized by a reduced number of color tones, resulting in abrupt transitions and an unnatural appearance. To perform the analysis, we compute and normalize the histograms for the RGB channels. Then, we count the number of bins below a certain predetermined threshold t_h and compare the resulting number against a threshold t. The score for this check is computed as:

$$\begin{gathered} C(h,t) = \begin{cases} 1 & \text{if } h \geq t \\ 0 & \text{otherwise} \end{cases} \\ S_{\text{NP}} = 1 - \frac{\sum_{i=0}^{255} C(h_i,t)}{t} \end{gathered} \tag{13}$$

The score is finally clipped between 0 and 1.

4) No pixelation: we determine whether an image is pixelated or not by computing the mean absolute difference between adjacent pixels of each row, thus obtaining a metric that presents a low value when in the presence of repetitive patterns and uniform blocks that are typical of pixelated images, and a high value when presented with an image

Subject Requirements	**ICAONet [13]**	**BioLab [14]**	**Correlance**	**Innovatrics**	**SDK 1**	**BioGaze (Ours)**
Head w/o coverings	0.175	0.260	0.040	-	0.280	**0.003**
Gaze in camera	0.445	0.538	0.201	0.272	0.257	**0.032**
Eyes open	0.125	0.500	**0.000**	**0.000**	0.203	0.016
No/light makeup	-	-	-	-	-	**0.002**
Neutral expression	0.339	0.272	0.109	0.110	0.158	**0.093**
No sunglasses	0.027	0.199	**0.000**	0.117	0.029	0.014
Frontal pose	0.234	-	-	-	0.374	**0.093**
Mean	*0.224*	*0.354*	*0.070*	*0.125*	*0.217*	***0.036***
Photographic Requirements						
Correct exposure	0.358	0.158	0.294	0.262	0.186	**0.105**
In focus photo	0.368	0.011	0.006	0.076	0.029	**0.000**
Correct saturation	0.321	0.053	0.421	-	**0.014**	0.041
Proper face dimension	-	-	0.010	-	0.010	-
Mean	*0.349*	*0.074*	*0.183*	*0.169*	*0.060*	***0.049***
Acquisition Requirements						
Uniform background	0.417	0.362	**0.114**	-	0.374	0.117
Uniform face lighting	0.376	-	0.101	0.286	0.249	**0.085**
No pixelation	0.494	0.312	-	-	0.032	**0.000**
No posterization	-	-	0.224	0.500	-	**0.008**
Mean	*0.429*	*0.337*	*0.146*	*0.393*	*0.218*	***0.053***
Global Mean	0.307	0.267	0.127	0.203	0.169	**0.044**

TABLE I: Performance comparison measured in Equal Error Rate (EER) of our proposed approach against various methods from the literature, *i.e.* ICAONet [13] and BioLab-ICAO-Check [14], as well as commercial off-the-shelf SDKs, namely Correlance, Innovatrics, and an anonymous SDK (referred to as SDK 1). Lower EER values indicate better performance.

with more variations and finer details. This value is then normalized between 0 and 1 by employing two empirical thresholds of 19 and 106.

5) Photo in focus: to determine whether a photo is in focus, we estimate its blurriness using the Laplacian operator, a widely used method for edge detection that captures intensity variations. Blurry images, characterized by minimal and gradual changes in intensity, result in a low Laplacian response due to the absence of sharp edges. Our approach involves converting the image to grayscale and applying the Laplacian operator. The sharpness is then quantified by calculating the maximum absolute value in the resulting gradient matrix. Finally, the score is obtained as follows:

$$S_{\text{PF}} = \sigma\left(s - t\right) \tag{14}$$

where σ is the sigmoid function, s is the sharpness computed as described above, and $t = 0.545$ is a predetermined threshold under which the photo is considered out of focus.

IV. Experiments

To evaluate the effectiveness of our proposed method, we perform a comparison against several other ISO/ICAO compliance verification tools, both commercial SDKs and state-of-the-art approaches from the literature.

For a standardized evaluation, we use the TONO dataset [6], a synthetic dataset specifically designed for ISO/ICAO compliance testing. This dataset is composed of approximately 4000 images, each with a singular element that violates the ISO/ICAO guidelines; this feature renders it an optimal candidate for our experimental evaluation, as it enables us to evaluate a single compliance check in isolation.

We report the performance for each method and requirement in terms of Equal Error Rate (EER), defined as the point at which the False Acceptance Rate (FAR) and False Rejection Rate (FRR) curves intersect [4]. As computing these metrics requires both positive and negative samples, we extend the TONO dataset by including synthetic ISO/ICAO-compliant images taken from the ONOT [11] dataset.

Our chosen competitors from the literature include ICAONet [13] and BioLab-ICAO-Check [14] (here shortened with BioLab); moreover, we also compare our results against commercial off-the-shelf SDKs from Correlance, Innovatrics, and another commercial SDK whose name cannot be published for Non-Disclosure Agreements (NDA), named SDK 1 in the following.

A. Implementation Details

Several checks employ scoring functions that require the use of predetermined thresholds. While some of them are directly derived from the ISO/ICAO standards, others are instead computed algorithmically on the training set provided in the BioLab-ICAO [30] dataset, which is composed of 571 images of real subjects, manually labeled by experts.

As the methods chosen for comparison may implement equivalent tests under different names (see Sect. II-B), we map those to the defects present in the TONO test set. For instance, the "unnatural skin tone" check present in BioLab-ICAO-Check and ICAONet is used to evaluate image saturation, "too light/too dark" assesses exposure, "roll/pitch/yaw" evaluates the subject's pose, and "mouth open" checks the facial expression. For Correlance, the "dynamic color" test is used to check for posterization, while "color control" measures the exposure, and "shadow presence" verifies the lighting uniformity. Similarly, we map the "contrast" and "specularity" checks provided by Innovatrics to respectively assess exposure and facial illumination.

B. Discussion of Results

Experimental results, detailed in Table I, include checks divided into three main categories: subject, photographic, and acquisition requirements. For each category, we report the mean EER across all implemented checks, along with the global mean EER to provide a synthetic representation of a method's overall performance.

Firstly, we observe that our proposed method is able to tackle almost all the requirements present in the dataset, while some competitors do not implement several fundamental checks in order to guarantee ISO/ICAO compliance. In particular, we note that several methods, both public and commercial SDKs, fail to verify the "frontal pose" requirement, which is critical to ensure the best performance of automatic face recognition systems.

Secondly, experimental results show that our proposed method is capable of outperforming all the selected competitors in the vast majority of requirements. Furthermore, these findings highlight significant improvements in certain checks compared to the state of the art; *e.g.* "gaze in camera" achieves a $6.2\times$ improvement, while "no posterization" achieves a $28\times$ improvement compared to the best result reported by any competitor. However, handling the "uniform background" requirement remains a complex challenge for all methods examined, as reflected in the consistently high EER values, indicating that achieving satisfactory performance in this task is particularly challenging.

For some commercial SDKs, a subset of the test images have been rejected without processing them, and they remain unevaluated. Unfortunately, due to the proprietary and closed-source nature of these solutions, the underlying reasons for such rejections cannot be determined precisely. In particular, the Correlance SDK rejected 3.7% of the images when evaluating the performance of the "no sunglasses" requirement. This behavior may be likely attributed to failures in the face detection process, potentially caused by the eyes not being visible through the dark lenses. Additionally, Innovatrics is not able to evaluate the "uniform background" check successfully in all images and therefore its performance for that requirement is not included.

To gain deeper insights into BioGaze's behavior, we conducted a visual analysis of the results and highlighted notable cases in Fig. 5. Specifically, the first row (Fig. 5a-5d) presents examples from the "mixed" set of TONO images, which lacks ground-truth labels and contains samples generally exhibiting multiple non-compliant characteristics. The selected cases are particularly interesting as they demonstrate instances where our SDK successfully identified these non-compliant traits, as confirmed through visual inspection

(a) Background, pose, lighting (b) Background, pose, lighting (c) Makeup (d) Expression, eyes, pose

(e) Expression, Gaze (f) Background (g) Shoulder (h) Compliant

Fig. 5: Visual examples of correct (first row) and wrong (second row) BioGaze compliance verification cases.

(details provided in the caption). The second row highlights some failure cases. Specifically, Fig. 5e presents a compliant image that did not pass the expression and gaze check. This is a borderline case, where even a human examiner might struggle to determine compliance. Fig. 5f shows another compliant image misclassified as non-compliant due to background uniformity issues, likely caused by shadows behind the head or shoulders. In Fig. 5g, a compliant image is incorrectly flagged as non-compliant due to the frontal shoulder check; in this instance, hair on the shoulder obscures symmetry evaluation between the left and right sides. Finally, Fig. 5h presents a non-compliant image due to a non-uniform background, yet it was incorrectly classified as compliant. Here, the background check failed because the inconsistencies in uniformity were minimal.

V. Conclusions and Future Works

In this paper, we have introduced BioGaze, a novel ISO/ICAO compliance verification tool designed to address the challenges associated with verifying facial image quality standards. By leveraging a combination of classical computer vision techniques and deep learning algorithms, BioGaze achieves state-of-the-art performance in classifying compliant images, demonstrating superior accuracy over existing public and commercial solutions. BioGaze's ability to integrate seamlessly into automated systems makes it ideal for deployment in high-throughput environments such as passport issuance centers. Additionally, the framework's open-source availability ensures transparency and accessibility, fostering further research and development in ISO/ICAO compliance verification. Several future works can be planned. Firstly, it would be desirable to develop a method that does not rely on any predefined and empirical threshold: this would enhance its generalization ability to new data, eliminating the need for a manual initial settings phase. Moreover, We plan to test the framework on the FVC-onGoing platform, but this requires a specific software implementation and fulfilling strict time constraints.

Ethical Impact Statement

BioGaze is a novel framework for ISO/ICAO compliance verification of facial images and then the study has been conducted without the collection of new data, nor direct interaction with partecipants. Moreover, BioGaze has been tested using publicly available synthetic datasets, which do not involve identifiable human data. The automated nature of BioGaze could introduce risks such as reinforcing biases inherent in the datasets used for the training of the models for which we downloaded the public weights. Ethnic, cultural, and religious variability in facial features may lead to inconsistent performance across demographic groups, potentially increasing discrimination or inequalities. On the other side, BioGaze introduces substantial benefits, including the potential to enhance the quality of document images, reduce the burden of manual inspection, and improve the consistency of biometric evaluations. It is important to note that the system is not intended to be used directly in border controls but only to enhance the visual quality of facial images. Moreover, by making the framework openly available, the research community gains access to a robust tool that can spur innovation and further advancements in facial recognition and compliance verification.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. 2010.

[2] A. Ahmadvand and M.-S. Moin. Estimating conformity of head yaw to the icao standard using a convolutional neural network. In *2018 9th International Symposium on Telecommunications (IST)*, pages 69–73, 2018.

[3] E. V. C. L. Borges, I. L. P. Andrezza, J. R. T. Marques, R. A. T. Mota, and J. J. B. Primo. Analysis of the eyes on face images for compliance with iso/icao requirements. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 173–179, 2016.

[4] G. Borghi, N. Di Domenico, A. Franco, M. Ferrara, and D. Maltoni. Revelio: A modular and effective framework for reproducible training and evaluation of morphing attack detectors. *IEEE Access*, 11:120419–120437, 2023.

[5] G. Borghi, A. Franco, N. Di Domenico, M. Ferrara, and D. Maltoni. V-mad: Video-based morphing attack detection in operational scenarios. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2024.

[6] G. Borghi, A. Franco, N. Di Domenico, and D. Maltoni. TONO: a Synthetic Dataset for Face Image Compliance to ISO/ICAO Standard. In *European Conference on Computer Vision*. Springer, 2004.

[7] K. W. Bowyer. Face recognition technology: security versus privacy. *IEEE Technology and society magazine*, 23(1):9–19, 2004.

[8] J. S. Dai. Euler–rodrigues formula variations, quaternion conjugation and intrinsic connections. *Mechanism and Machine Theory*, 92:144–152, 2015.

[9] K. Delac and M. Grgic. A survey of biometric recognition methods. In *Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine*, pages 184–193. IEEE, 2004.

[10] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Face restoration for morphed images retouching. In *2024 12th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2024.

[11] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Onot: a high-quality icao-compliant synthetic mugshot dataset. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10, 2024.

[12] W. Dong and Z. Shisheng. Color image recognition method based on the prewitt operator. In *2008 International Conference on Computer Science and Software Engineering*, volume 6, pages 170–173. IEEE, 2008.

[13] A. G. d. A. e Silva, H. M. Gomes, and L. V. Batista. A collaborative deep multitask learning network for face image compliance to iso/iec 19794-5 standard. *Expert Systems with Applications*, 198:116756, 2022.

[14] M. Ferrara, A. Franco, D. Maio, and D. Maltoni. Face image conformance to iso/icao standards in machine readable travel documents. *IEEE Transactions on Information Forensics and Security*, 7(4):1204–1213, 2012.

[15] M. Ferrara, A. Franco, and D. Maltoni. Fingerprint scanner focusing estimation by top sharpening index. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 223–228. IEEE, 2007.

[16] I. Goodfellow. Deep learning, 2016.

[17] C. Guerra, J. Marcos, and N. Gonçalves. Automatic validation of icao compliance regarding head coverings: An inclusive approach concerning religious circumstances. In *2023 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–4, 2023.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] International Civil Aviation Organization (ICAO). Machine readable travel documents. part 11: Security mechanisms for MRTDs. Standard, International Civil Aviation Organization, 2015.

[20] International Standards Organization. ISO/IEC 14496-2 — Information technology — Coding of audio-visual objects — Part 2: Visual. Standard, International Organization for Standardization, 2004.

[21] International Standards Organization. ISO/IEC 19794-5 — Information technology — Biometric data interchange formats — Part 5: Face image data. Standard, International Organization for Standardization, 2011.

[22] International Standards Organization. ISO/IEC 39794-5 — Information technology — Extensible biometric data interchange formats — Part 5: Face image data. Standard, International Organization for Standardization, 2019.

[23] International Standards Organization. ISO/IEC 29794-5 — Information technology — Biometric sample quality — Part 5: Face image data. Standard, International Organization for Standardization, under development.

[24] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann. Real-time facial surface geometry from monocular video on mobile gpus. *arXiv preprint arXiv:1907.06724*, 2019.

[25] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[26] D. E. King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.

[27] C.-H. Lee, Z. Liu, L. Wu, and P. Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5549–5558, 2020.

[28] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[29] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, et al. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*, 2019.

[30] D. Maltoni, A. Franco, M. Ferrara, D. Maio, and A. Nardelli. Biolab-icao: A new benchmark to evaluate applications assessing face image compliance to iso/iec 19794-5 standard. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 41–44. IEEE, 2009.

[31] E. Marchand, H. Uchiyama, and F. Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12):2633–2651, 2015.

[32] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[33] A. Mazandarani, P. M. F. Amaral, P. da Fonseca Pinto, and S. J. H. Shamoushaki. Deep learning-based automated detection of inappropriate face image attributes for id documents. In L. M. Camarinha-Matos, P. Ferreira, and G. Brito, editors, *Technological Innovation for Applied AI Systems*, pages 243–253, Cham, 2021. Springer International Publishing.

[34] J. Merkle, C. Rathgeb, B. Herdeanu, B. Tams, D.-P. Lou, A. Dörsch, M. Schaubert, J. Dehen, L. Chen, X. Yin, D. Huang, A. Stratmann, M. Ginzler, M. Grimmer, and C. Busch. Open source face image quality (ofiq): Implementation and evaluation of algorithms. Technical report, Federal Office for Information Security, 2024.

[35] R. L. Parente, L. V. Batista, I. L. P. Andrezza, E. V. C. L. Borges, and R. A. T. Mota. Assessing facial image accordance to iso/icao requirements. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 180–187, 2016.
[36] L. Pham, T. H. Vu, and T. A. Tran. Facial expression recognition using residual masking network. In *2020 25Th international conference on pattern recognition (ICPR)*, pages 4513–4519. IEEE, 2021.
[37] D. Reis, J. Kupec, J. Hong, and A. Daoudi. Real-time flying object detection with yolov8. *arXiv preprint arXiv:2305.09972*, 2023.
[38] A. N.-S. M. Sharifi. Facial images quality assessment based on iso/icao standard compliance estimation by hmax model. *Journal of Information Systems and Telecommunication (JIST)*, 7(3), 2020.
[39] A. Wolf. Icao: Portrait quality (reference facial images for mrtd), version 1.0. standard. *International Civil Aviation Organization*, 2018.
[40] J. Yang, P. J. Grother, M. L. Ngan, K. Hanaoka, and A. Hom. Face analysis technology evaluation (fate) part 11: Face image quality vector assessment: Specific image defect detection. 2023.
[41] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 325–341, 2018.
[42] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges. Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 365–381. Springer, 2020.
[43] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458, 2003.