IJCB 2026 TUTORIAL

When Two Faces Become One

Part II — Detecting, Evaluating, and Deploying Morphing Attack Countermeasures

Marija Ivanovska

UNIVERSITY OF LJUBLJANA | FE Faculty of Electrical Engineering

IEEE International Joint Conference on Biometrics

Rome, Italy · September 1, 2026

Roadmap for Part II

SESSION 3

Detection methods

Morphing evidence,
S-MAD / D-MAD / V-MAD,
detector families, generalization

DEMO

S-MAD
&
D-MAD

Morphing score,
algorithm
robustness

SESSION 4

Evaluation

Metrics, protocols,
independent benchmarks

Learning path: evidence → methods → limitations → evaluation → deployment

Why Morphing Attack Detection (MAD) Matters

Document issuance

A morphed portrait is submitted during application.

Border / gate verification

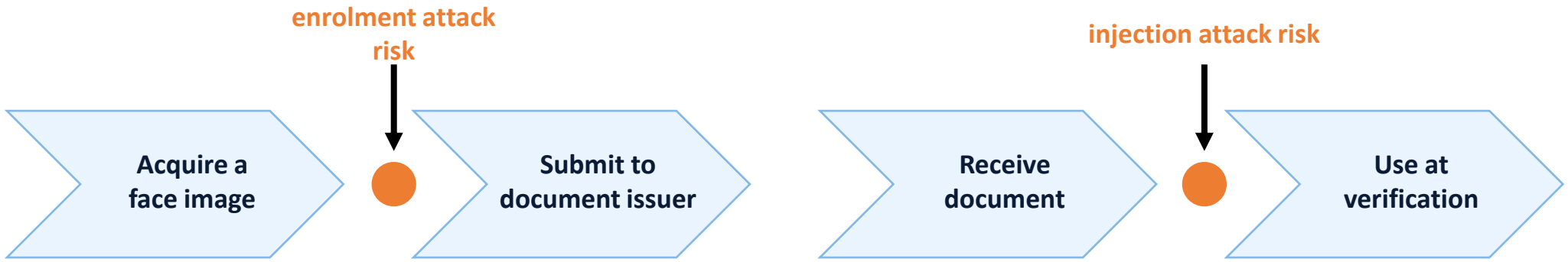
The document image may match more than one person.

Digital onboarding

Remote identity proofing relies on submitted face images.

Ferrara et.al, The Magic Passport, IJCB, 2014

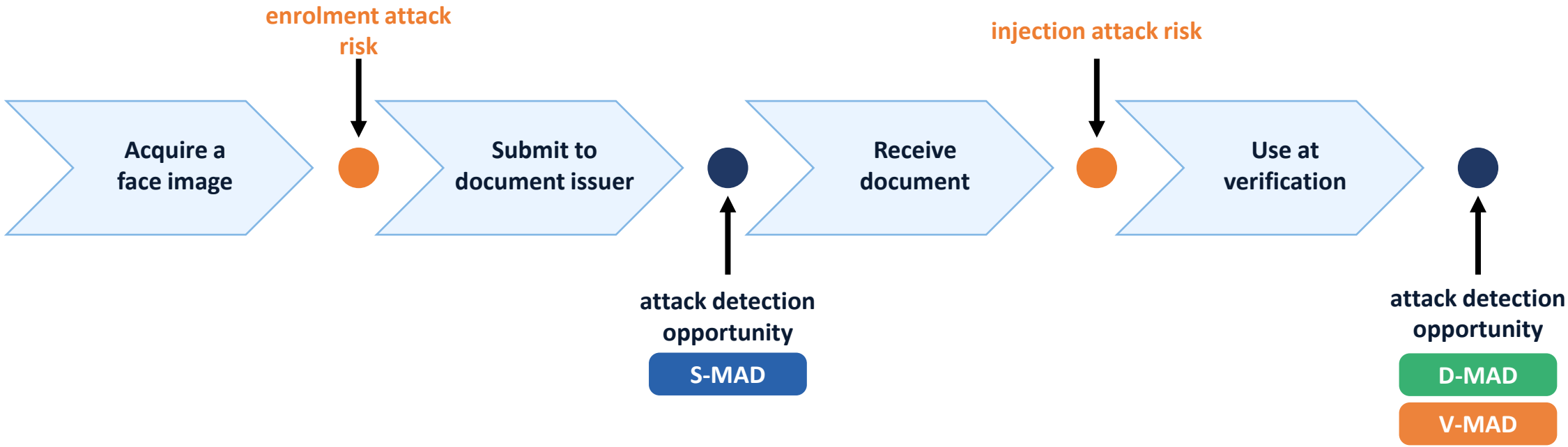
Morphing Attack Pipeline



Attacks can be introduced before document issuance or injected into a valid document.

Scherhag et al., Face Recognition Systems under Morphing Attacks: A Survey. IEEE Access, 2019

Morphing Attack Pipeline



Scherhag et al., Face Recognition Systems under Morphing Attacks: A Survey. IEEE Access, 2019

Detection Starts With the Evidence That is Available

S-MAD

INPUT

One suspect image

EVIDENCE

Artifacts in the image space and the frequency space

USE CASE

Document issuance, remote enrollment

D-MAD

INPUT

Suspect image + trusted reference / live image

EVIDENCE

Cross-image identity and feature inconsistencies

USE CASE

ID verification, border control

V-MAD

INPUT

Suspect image + sequence of live frames

EVIDENCE

Multi-frame identity features

USE CASE

Operational gates with video

Detector design should follow the evidence available - not the current model fashion.

Scherhag et al., Face Recognition Systems under Morphing Attacks: A Survey. IEEE Access, 2019

S-MAD: Why is S-MAD Difficult

1 **No reference image**

No live/reference image is available for identity comparison.

2 **Evolving attack generators**

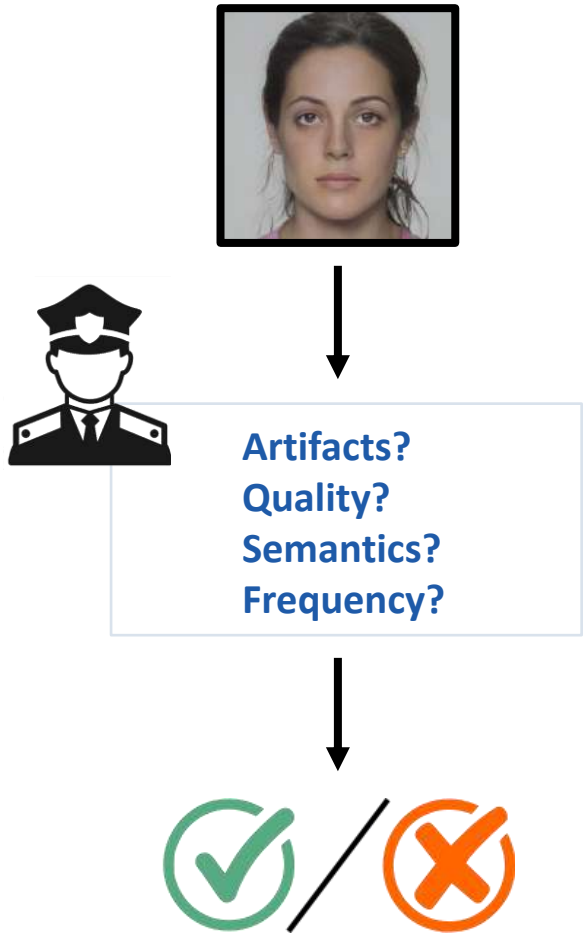
Different morph generators leave different traces, and these traces evolve as generation methods improve.

3 **Operational image variability**

Compression, resizing, print-scan processes, lighting, pose, and cropping can mimic or hide morphing traces.

4 **False alarms are costly**

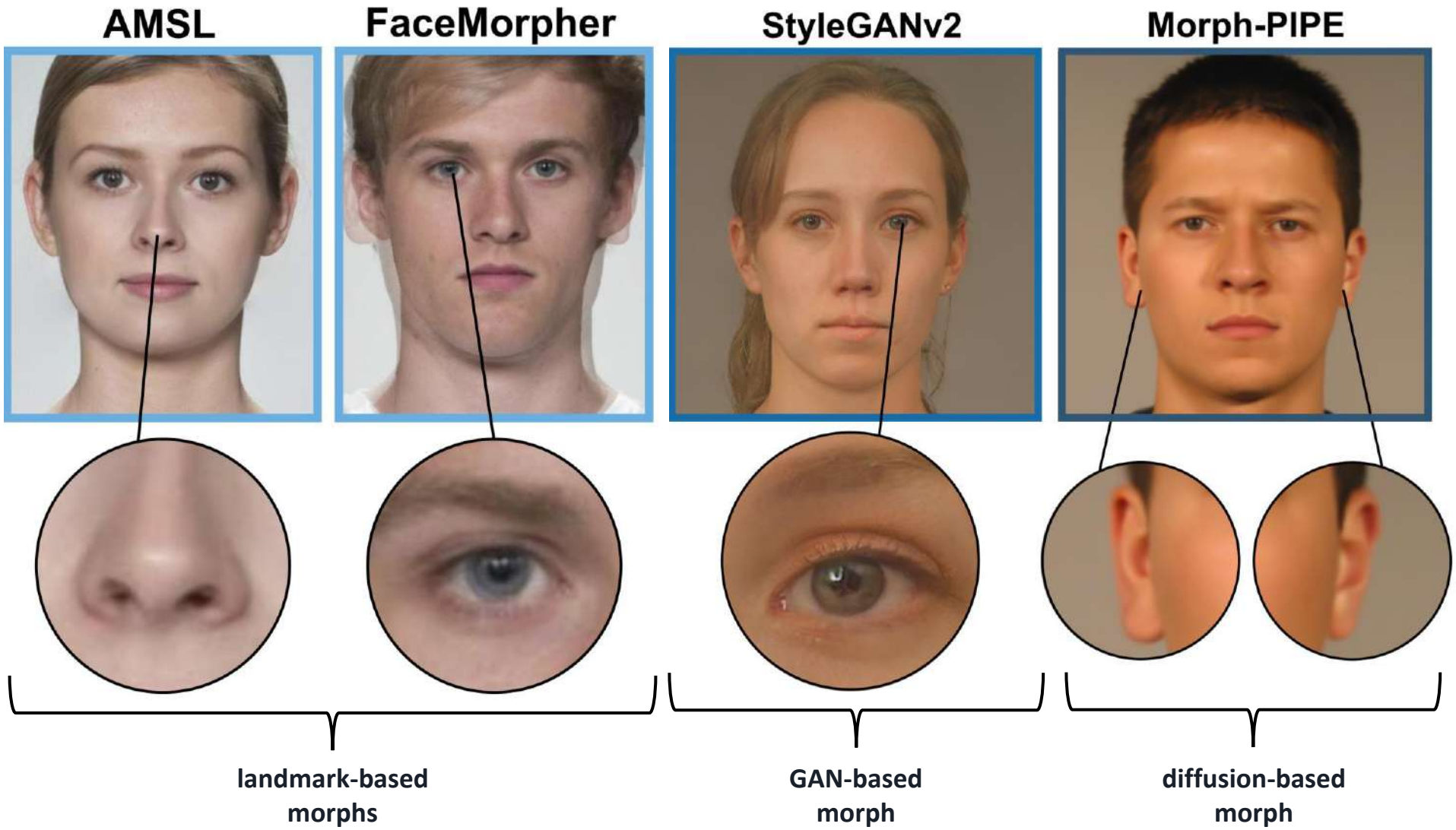
In document issuance or border-control settings, rejecting bona fide applicants creates workload, delays, and trust issues.



S-MAD is an open-set generalization problem under constrained evidence.

S-MAD: What Morphing Artifacts Look Like

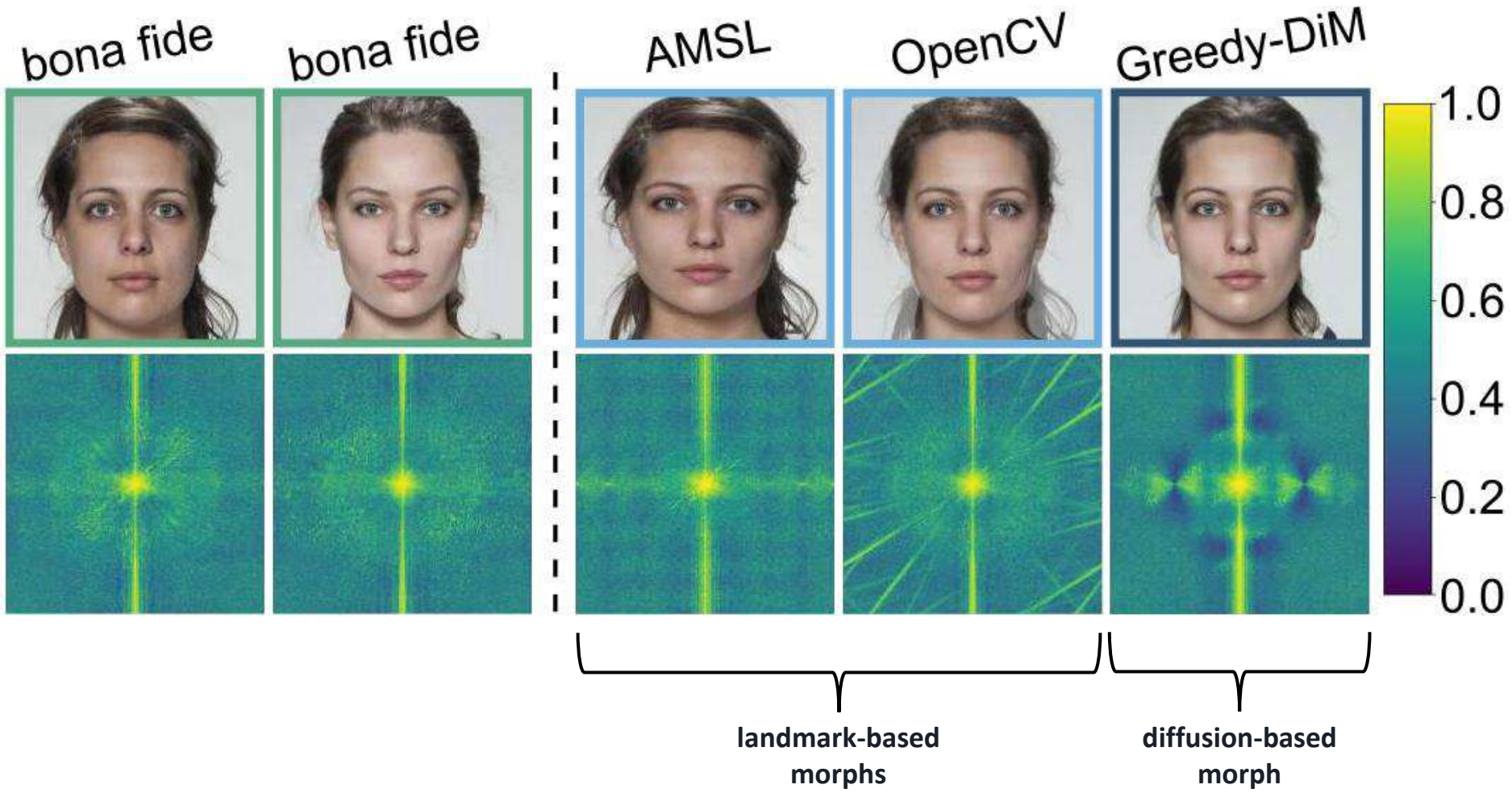
- Misaligned face parts
- Double contours
- Ghosting
- Over-smoothed skin
- Irregular pupils and irises
- Mismatched ears
- …



Ivanovska et al., SelfMAD++; Information Fusion 2026

S-MAD: What Morphing Artifacts Look Like

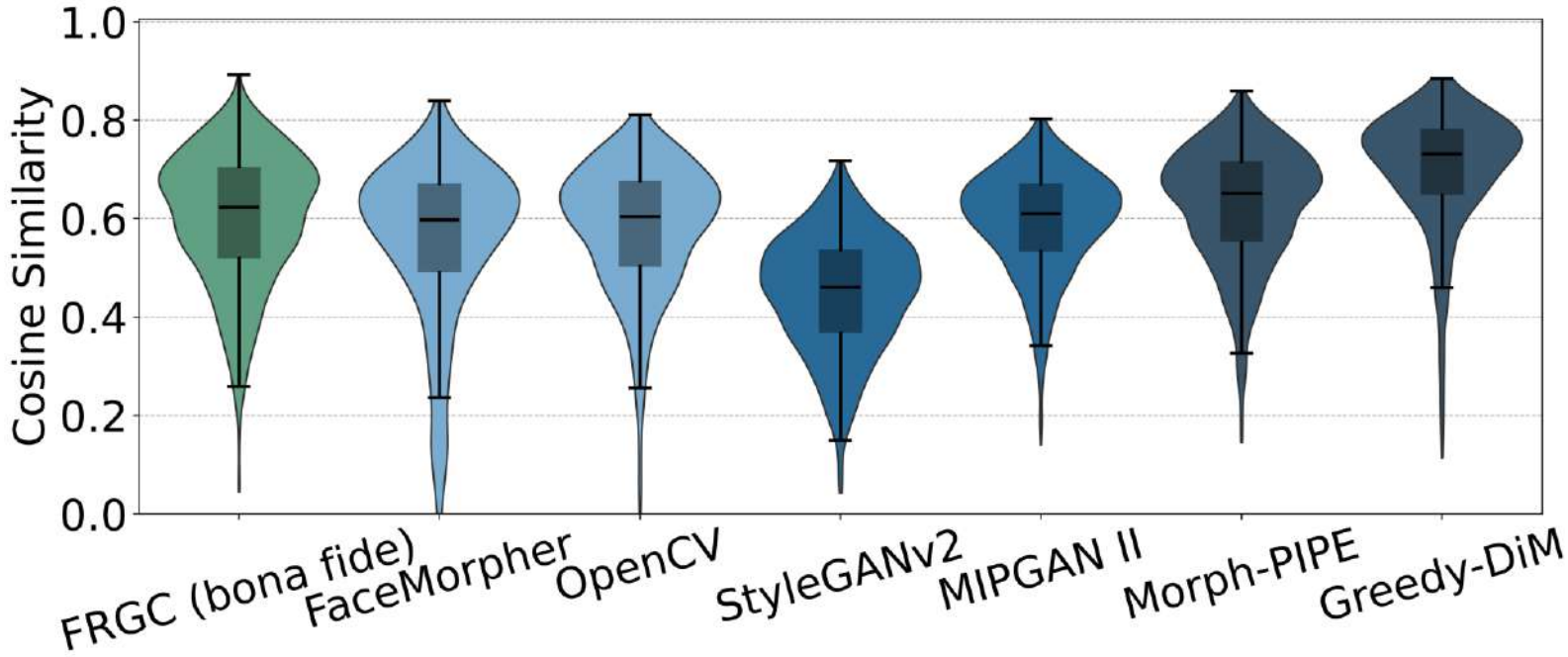
- …
- Suppressed or boosted high-frequency bands



Ivanovska et al., SelfMAD++; Information Fusion 2026

S-MAD: What Morphing Artifacts Look Like

Semantic consistency across face halves, measured by cosine similarity.



Bona fide and landmark-based morphs show natural semantic asymmetry, whereas GAN- and diffusion-based morphs may introduce **feature incoherence** or **unnaturally high semantic symmetry**.

Ivanovska et al., SelfMAD++; Information Fusion 2026

S-MAD Pipeline: From Evidence to a Decision

1 Pre-process
detect face, crop, align, normalize

2 Extract evidence
spatial artifacts, frequency artifacts, semantic irregularities

3 Produce a score
a ranking signal, not necessarily a probability

4 Apply a threshold
selected for a targeted risks

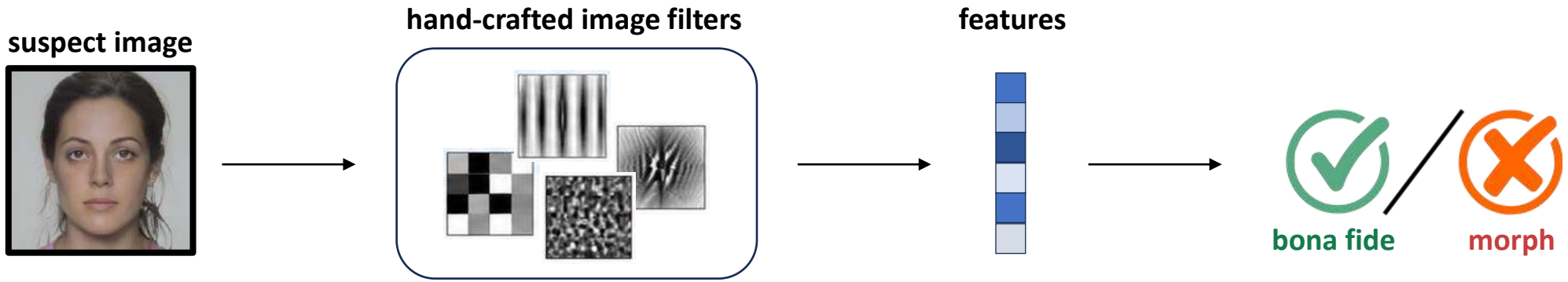
5 Take an action
accept/reject/ escalate to human review

- **Hand-crafted forensics** — analysis of texture, PRNU/noise, reflections etc.
- **Discriminative CNN / ViT** — learned local, global, spatial, and frequency cues
- **Quality / anomaly models** — deviation from bona fide statistics
- **Foundation / multimodal models** — transferable representations and semantic alignment

The score is useful only when combined with a predefined protocol, threshold, and action policy.

S-MAD: Hand-crafted forensics

- ***Texture descriptors***: LBP (Local Binary Patterns), BSIF (Binarized Statistical Image Features), HOG (Histogram of Oriented Gradients)
- ***Sensor-related evidence***: PRNU (Photo-Response Non-Uniformity), SPN (Sensor Pattern Noise), residual noise
- ***Geometry descriptors:*** landmark distances and ratios, shape inconsistencies, edge/contour statistics
- ***Frequency / forensic descriptors:*** *DCT/FFT/wavelet statistics, compression traces, reflection inconsistencies*



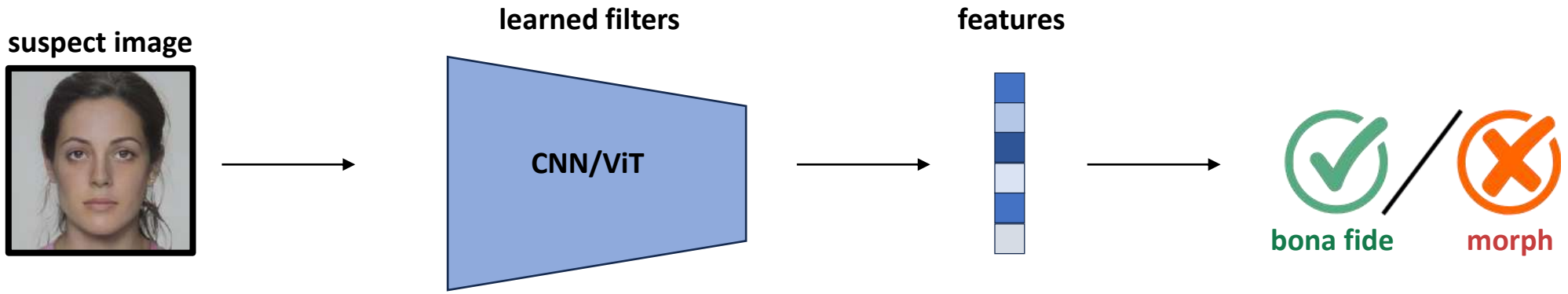
+ Data-efficient and often inspectable

– Fragile under post-processing and new generators

Raghavendra et al., Detecting Morphed Face Images, 2016; Debiasi et al., Detection of Face Morphing Attacks Based on PRNU Analysis, 2018; Ferrara et al., Detecting Morphing Attacks through Face Geometry Features, 2020

S-MAD: Discriminative CNN / ViT

- Learn **discriminative representations** of bona fide vs. morphed faces from labeled training data
- Capture **local evidence** around facial regions and fine-grained blending/texture artifacts
- Capture **global evidence** through face-wide consistency, spectral cues, and long-range ViT interactions



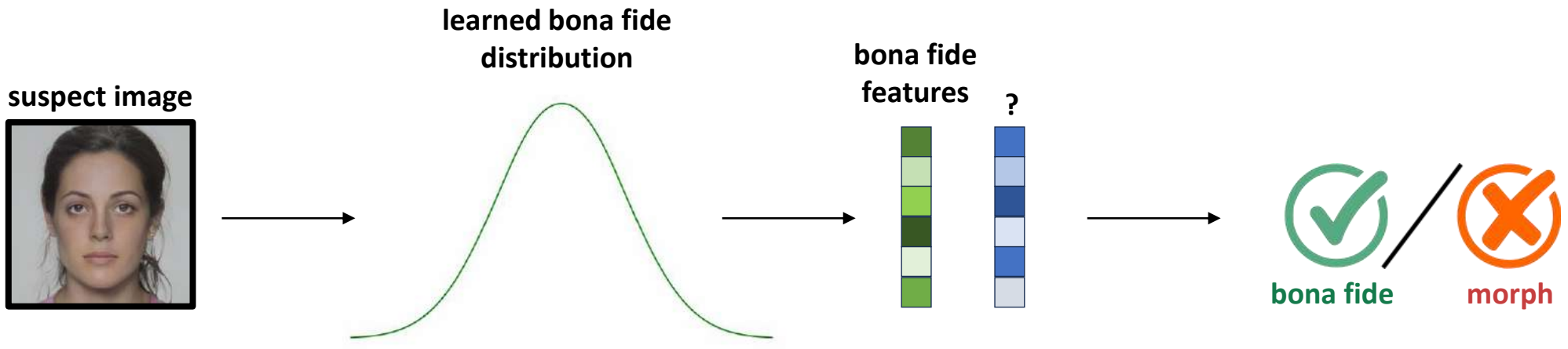
+ High capacity and very strong closed-set accuracy

– Often learn dataset or generator shortcuts

Damer et al., PW-MAD: Pixel-Wise Supervision for Generalized Face Morphing Attack Detection, 2021; Zhang et al., Generalized Single-Image-Based Morphing Attack Detection Using Deep Representations from Vision Transformer, 2024

S-MAD: Quality / anomaly models

- Learn representations of bona fide faces without explicitly modeling individual morphing attacks
- Capture deviations from bona fide appearance through quality cues, reconstruction residuals, density estimates, or self-supervised features



+ avoid attack-specific training

– May confuse low-quality bona fide images with attacks

Fu et al, Face Morphing Attacks and Face Image Quality: The Effect of Morphing and the Unsupervised Attack Detection by Quality, 2022;
Ivanovska et al., Face Morphing Attack Detection with Denoising Diffusion Probabilistic Models, 2023

S-MAD: Foundation / multimodal models

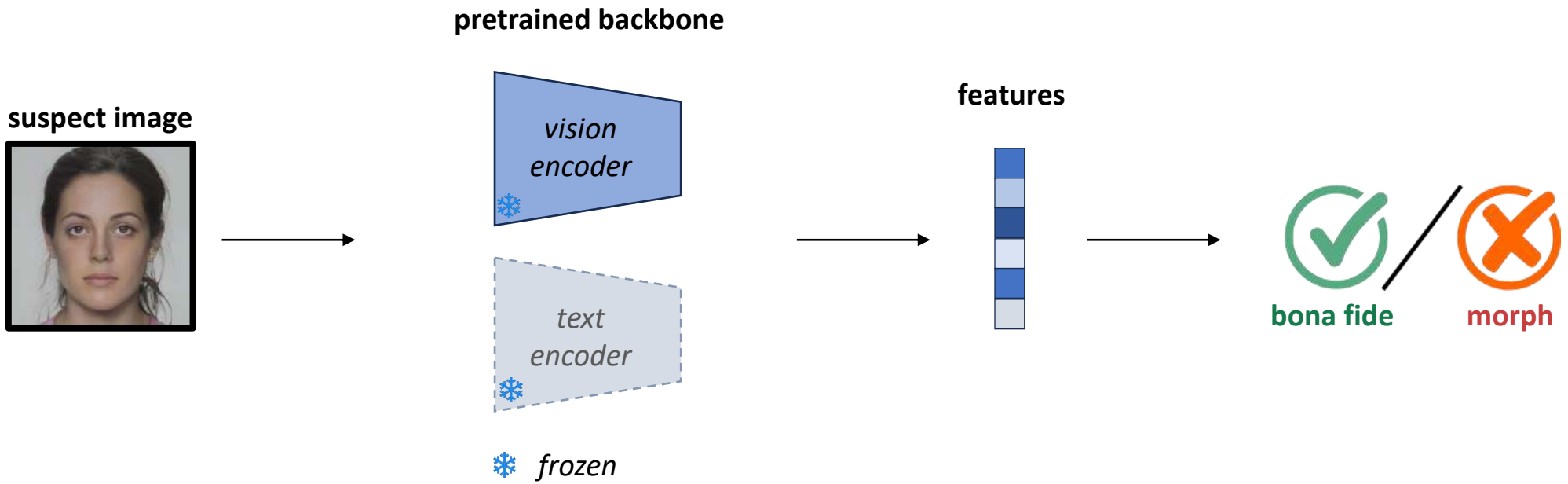
Foundation models encode broad prior knowledge. Common usage for MAD:

- **Zero-shot inference**
- **Feature extraction**
- **Parameter-efficient adaptation**
- **Interpretable Multimodal Reasoning**

S-MAD: Foundation / multimodal models

Derive a morph score or decision in a zero-shot setting using pretrained representations or multimodal alignment.

Zero-shot inference | Feature extraction | Parameter-efficient adaptation | Interpretable Multimodal Reasoning



+ No downstream task training

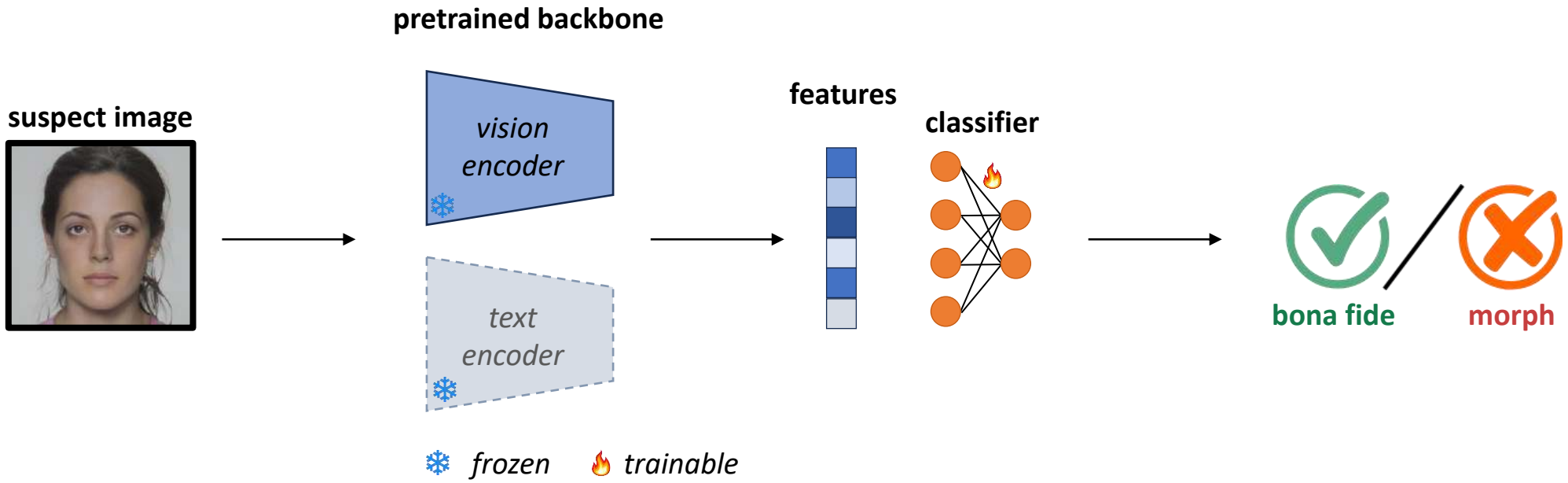
– Score calibration, prompt sensitivity, weak causal grounding

Patwardhan et al, Empowering Morphing Attack Detection Using Interpretable Image-Text Foundation Model, CVIP, 2024

S-MAD: Foundation / multimodal models

Extract pretrained vision-language features and train a lightweight classifier on top.

Zero-shot inference | **Feature extraction** | Parameter-efficient adaptation | Interpretable Multimodal Reasoning



+ Data-efficient and highly reproducible

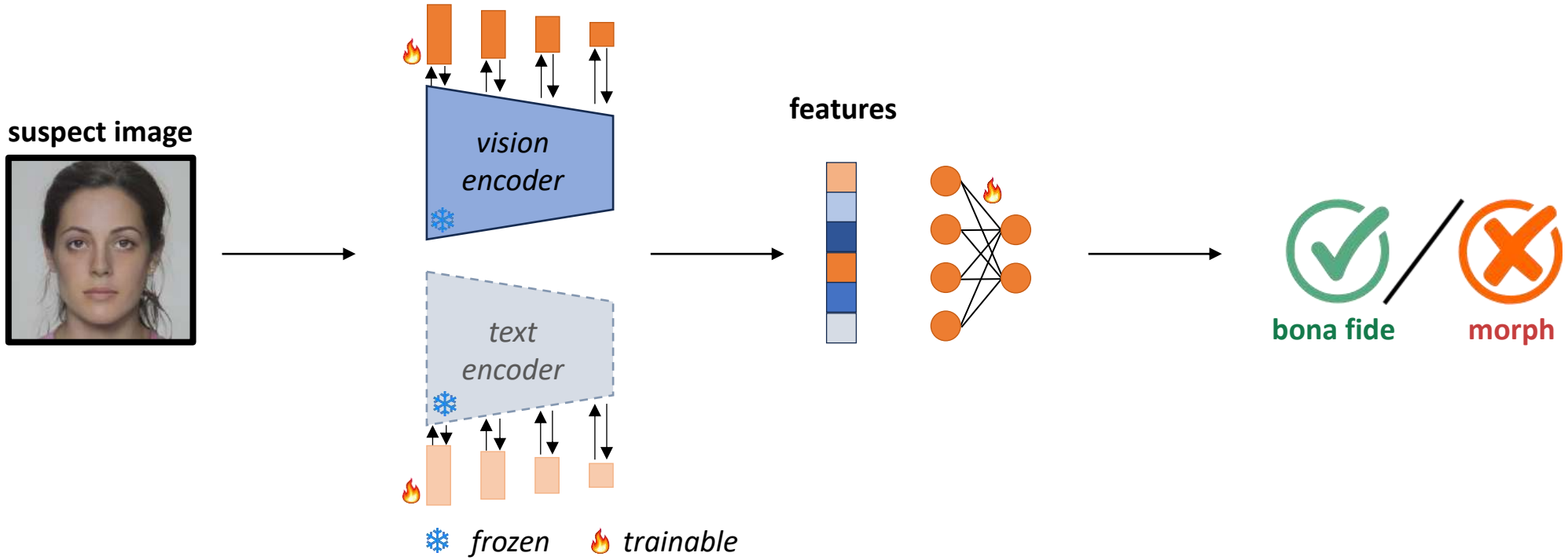
– Sensitive to feature-layer selection and domain shift

Caldeira et al., MADation: Face Morphing Attack Detection with Foundation Models, WACVW, 2025

S-MAD: Foundation / multimodal models

Downstream task adaptation using MAD data.

Zero-shot inference | Feature extraction | **Parameter-efficient adaptation** | Interpretable Multimodal Reasoning



+ Low trainable parameter count

– Risk of shortcut learning

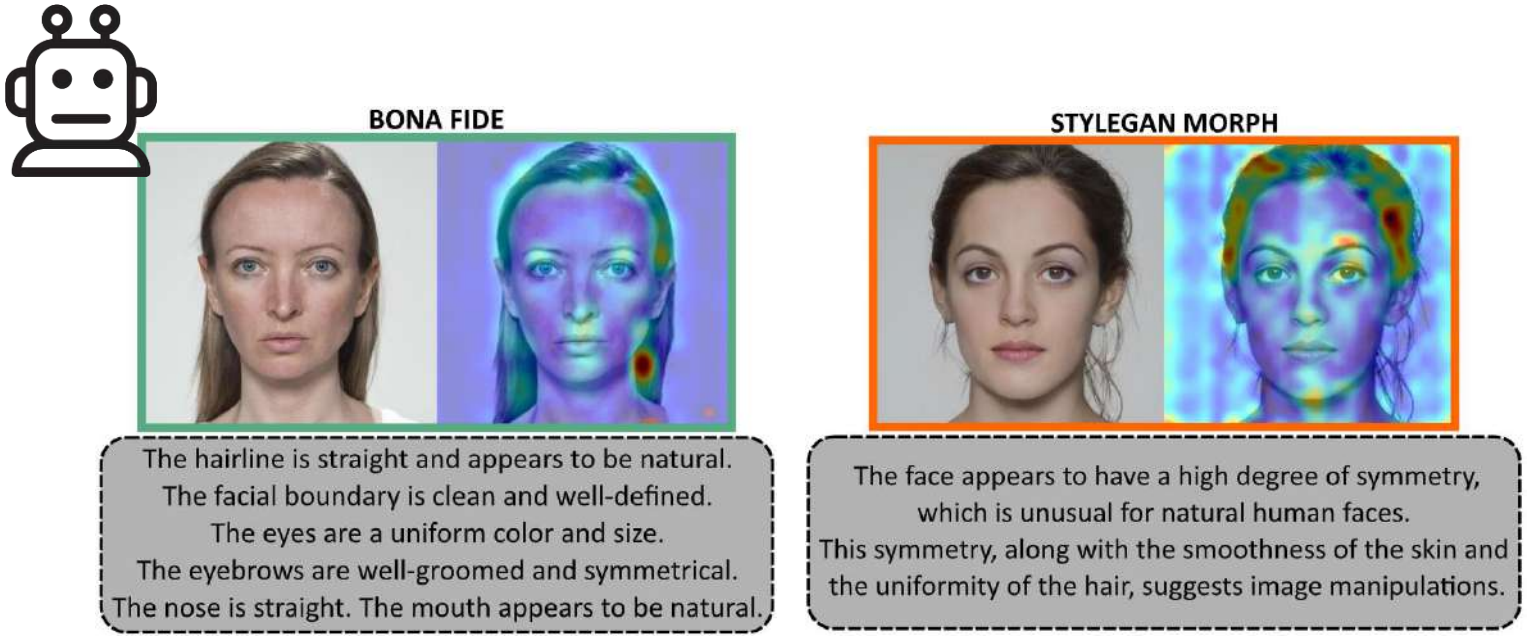
Ivanovska et al., SelfMAD++; Information Fusion 2026

S-MAD: Foundation / multimodal models

Perform human-interpretable detection or reasoning about morphing evidence.

Zero-shot inference | Feature extraction | Parameter-efficient adaptation | **Interpretable Multimodal Reasoning**

Is this image a face morphing attack? Describe the visual cues supporting your decision.



+ Human-interpretable detection and explanations

– Reasoning and explanations may not be well-grounded or faithful

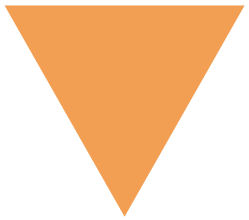
Ivanovska et al., Emergent Morphing Attack Detection in Open Multi-modal Large Language Models, https://arxiv.org/abs/2602.15461, 2026
Narayan et al., FaceXBench: Evaluating Multimodal LLMs on Face Understanding; T-BIOM 2026

S-MAD: Generalization as a Persisting Limitation

closed-set success ≠ operational robustness

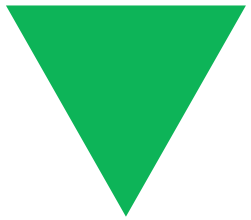
Dataset bias

Camera characteristics, demographics, background, cropping cues can dominate the decision.



Attack bias

Detectors often learn cues of specific morph generator(s), not the concept of morphing.



Quality bias

Compression, blur, resizing, or image enhancement may hide morphing evidence.

S-MAD: Demo

UBO-SMAD-R3

- **Inception-ResNet V1** trained for single-image MAD
- Trained in supervised manner using multiple morph generators and source datasets
- One of the SOTA S-MAD models on on the sequestered FVC-ongoing benchmarks
- *Input*: single face image
- *Output*: continuous morph score

What we will test

1 — Detection
Bona fide → landmark morph → GAN morph → diffusion morph

2 — Visualization
Grad-CAM: *where do neural activations happen?*

3 — Robustness
Same bona fide face → blur / downsampling → *does the score remain stable?*

Borghi et al., Revelio: A Modular and Effective Framework for Reproducible Training and Evaluation of Morphing Attack Detectors, IEEE Access, 2023

S-MAD: The Training-Data Problem

- ***Attack diversity***: What attacks do we train on?

  Morphing techniques continuously evolve: landmark-based, GAN-based, diffusion-based, retouched, post-processed, etc.

- ***Data quality***: Under what conditions are training samples observed?

  Resolution, compression, blur, print–scan, recapture, enhancement, and acquisition pipelines alter or suppress forensic evidence.

- ***Privacy & availability***: Where do training faces come from?

  Training MAD requires large collections of biometric face images. Such data are often privacy-sensitive, difficult to collect, share, and redistribute.

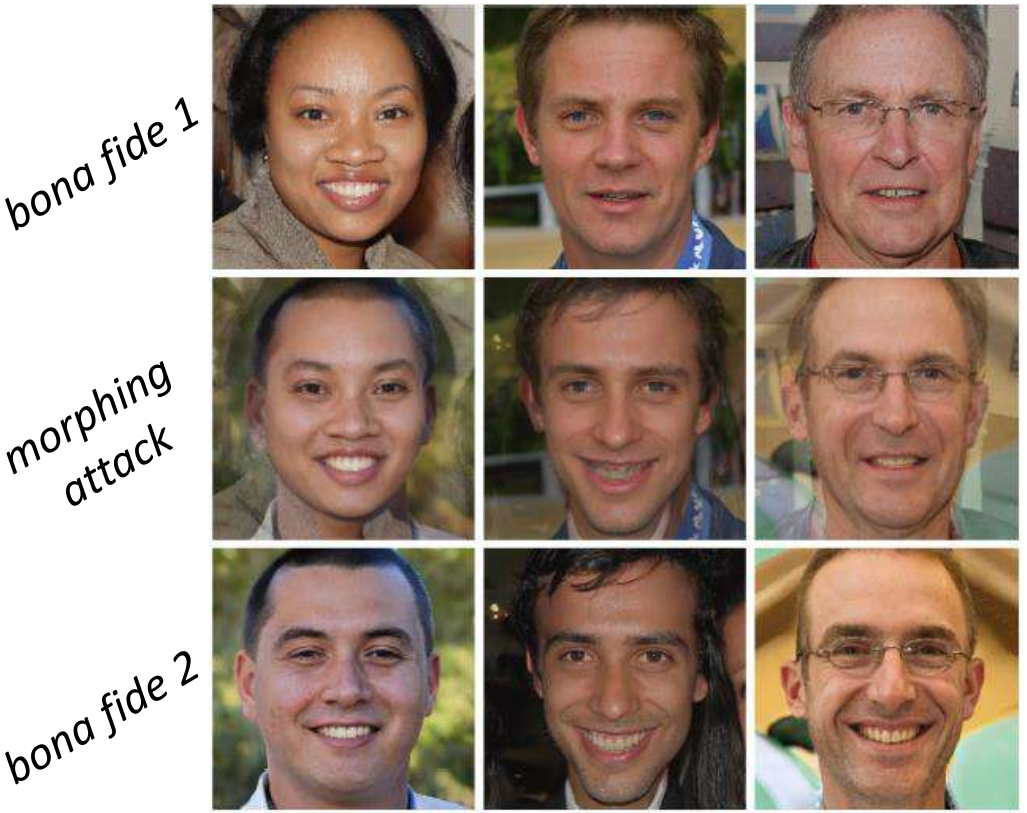
Can we generate the training signal differently?

S-MAD: Synthetic Data for Privacy-Aware MAD Training

IJCB 2022 Competition SYN-MAD: The Controlled Experiment

Can MAD be trained exclusively on synthetic identities?

- Training restricted to **synthetic data only (SMDD)**
- **25,000 synthetic bona fide images**
- **15,000 synthetic (OpenCV) morphing attacks**
- Evaluation performed on authentic/real biometric data with attack variation



> ***Key result:*** synthetic-only training is feasible, but synthetic-to-real generalization remains the challenge.

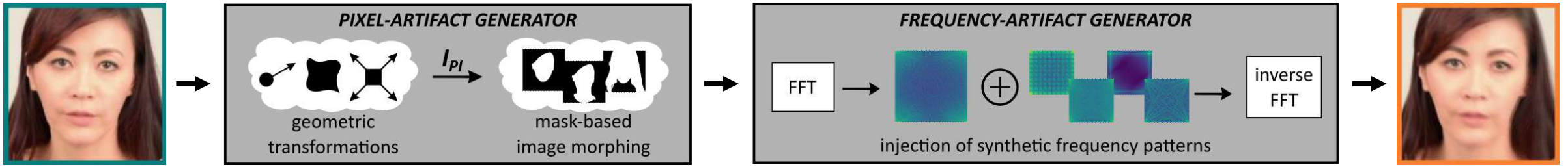
More recent synthetic datasets:

- **MONOT** (2024)- 15k morphs, six morphing algorithms, ISO/ICAO-oriented
- **SynMorph** (2025) - more than 100,000 morphs
- **FLUXSynID** (2025) - synthetic identity source for MAD

Huber et al., SYN-MAD 2022: Competition on Face Morphing Attack Detection Based on Privacy-aware Synthetic Training Data, IJCB, 2022

S-MAD: Creating Morph-like Evidence Without Real Morphing Techniques

SelfMAD is an augmentation technique that simulates diverse artifact patterns from bona fide images.



bona fide

Pixel-space artifact simulation

warping · blending · ghosting

Frequency-space artifact simulation

spectral perturbations · hidden traces

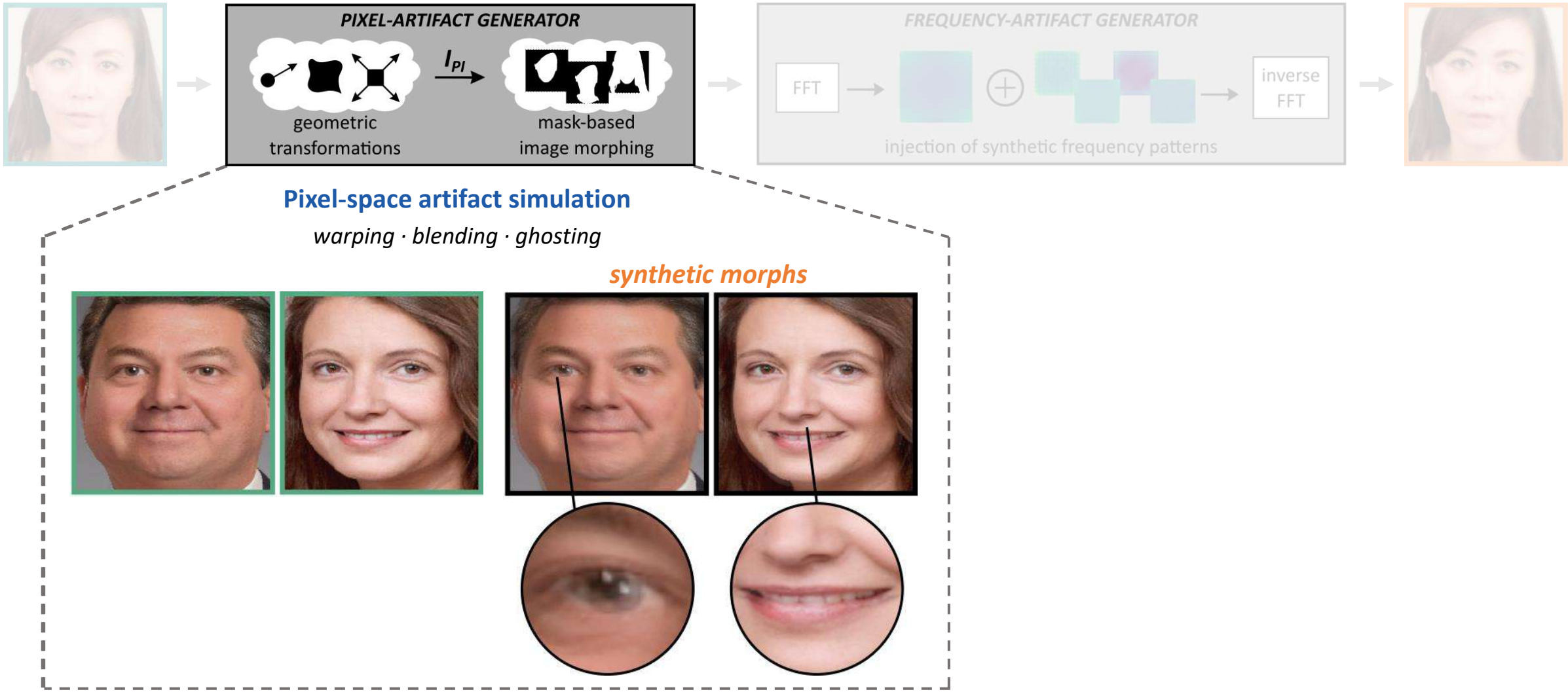
synthetic morph

Controlled artifact diversity helps detectors better learn morphing evidence, instead of dataset-, quality-, or generator-specific shortcuts.

Ivanovska et al., SelfMAD: Enhancing Generalization and Robustness in MAD via Self-Supervised Learning, FG, 2025

S-MAD: Creating Morph-like Evidence Without Real Morphing Techniques

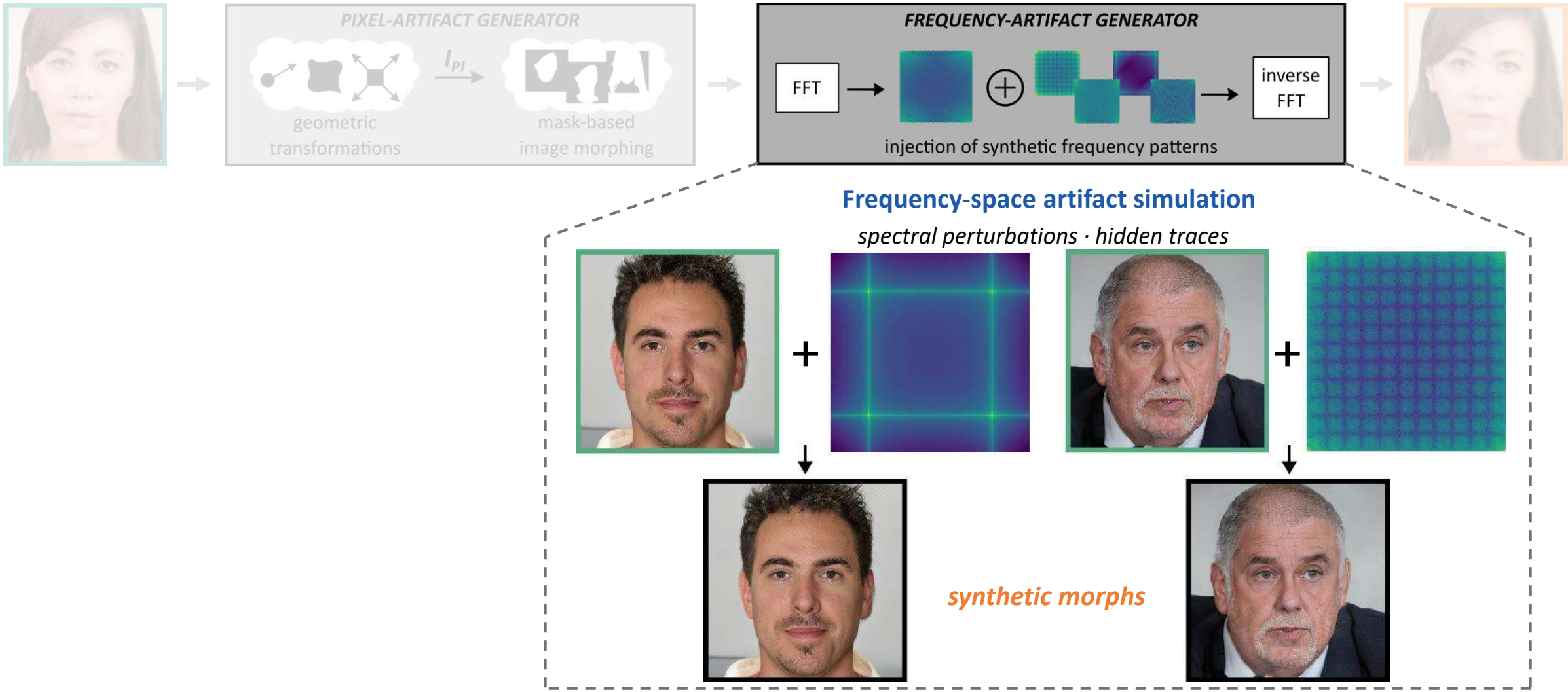
SelfMAD is an augmentation technique that simulates diverse artifact patterns from bona fide images.



Ivanovska et al., SelfMAD: Enhancing Generalization and Robustness in MAD via Self-Supervised Learning, FG, 2025

S-MAD: Creating Morph-like Evidence Without Real Morphing Techniques

SelfMAD is an augmentation technique that simulates diverse artifact patterns from bona fide images.



D-MAD: What Changes When a Reference Exists?

S-MAD	D-MAD
One suspect image	**Suspect image + trusted reference/live capture**
Looks for manipulation evidence	**Looks for cross-image inconsistencies**
Useful at enrolment/application stage	**Useful for verification, border control, or identity proofing**

The trusted reference introduces identity evidence that is fundamentally unavailable to S-MAD.
Caveat: D-MAD depends on the reference being genuinely trusted and representative.

D-MAD: Different Ways to Exploit the Image Pair

D-MAD transforms the problem from single-image artifact evidence into *comparative representation analysis*.

1 **Identity embedding comparison: global identity discrepancy**

Compare face-recognition representations of document and live image.

2 **Pairwise local comparison: cross-image inconsistencies**

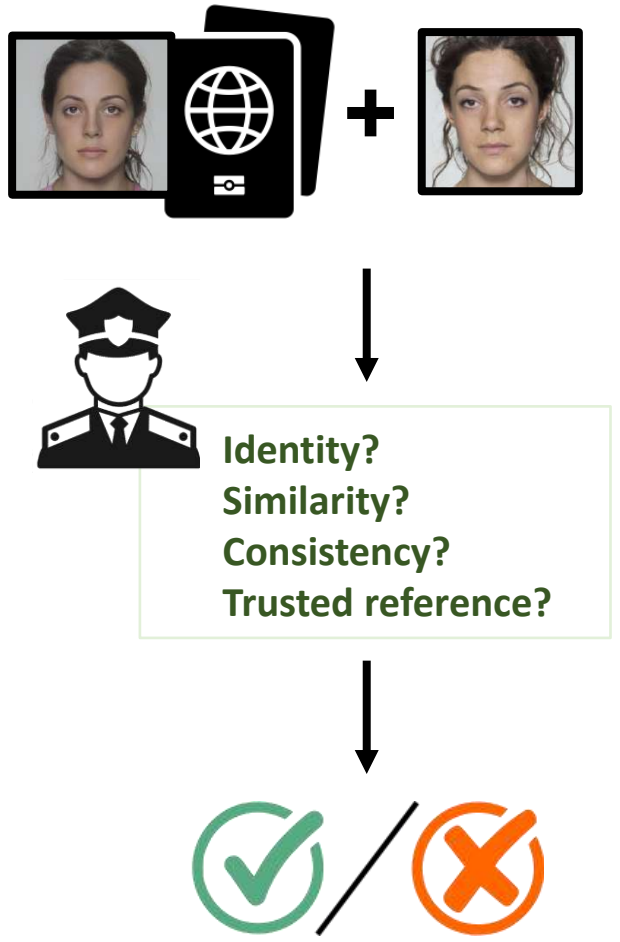
Learn where and how the two representations differ.

3 **Forensic + identity analysis: combined artifact and identity evidence**

Combine differential identity evidence with S-MAD-like manipulation evidence.

4 **Demorphing: residual hidden identity**

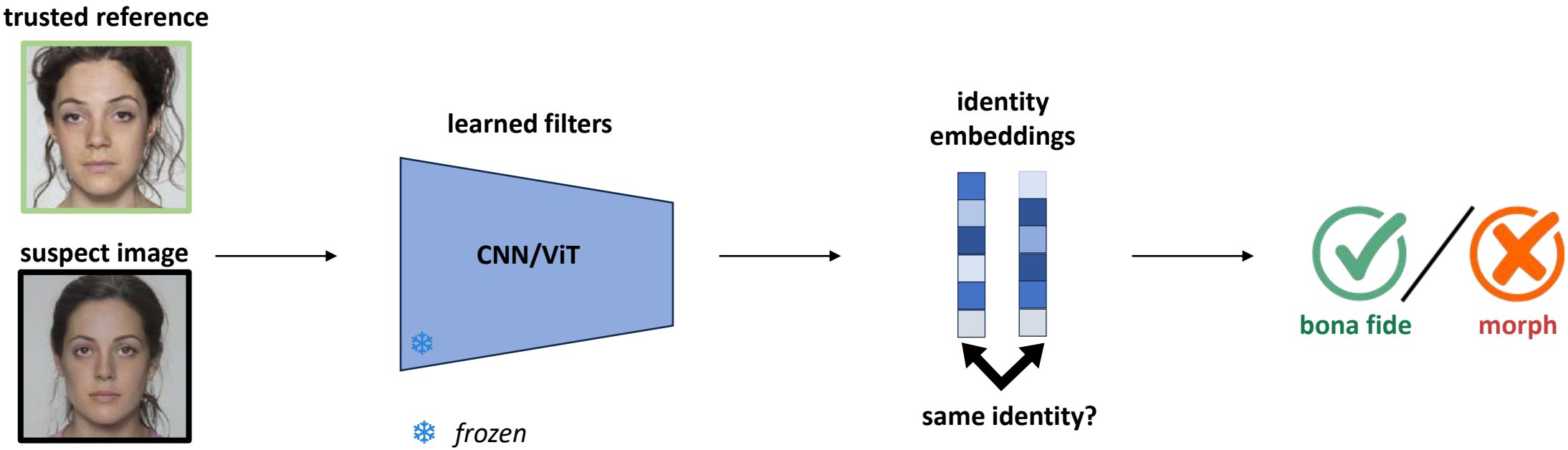
Use the trusted image to estimate/remove one identity contribution and inspect what remains.



All D-MAD methods exploit the trusted reference, but they differ in what representation is compared and how the difference is modeled.

D-MAD: Identity Embedding Comparison

- **Extract identity embeddings** from document and trusted live/reference images.
- Construct a **differential representation**, e.g. element-wise embedding difference.
- Learn whether the discrepancy resembles normal intra-person variation or **morph-induced identity inconsistency.**
- **D-MAD is not the same as face matching**, as the method learns whether the cross-image discrepancy is characteristic of morphing.



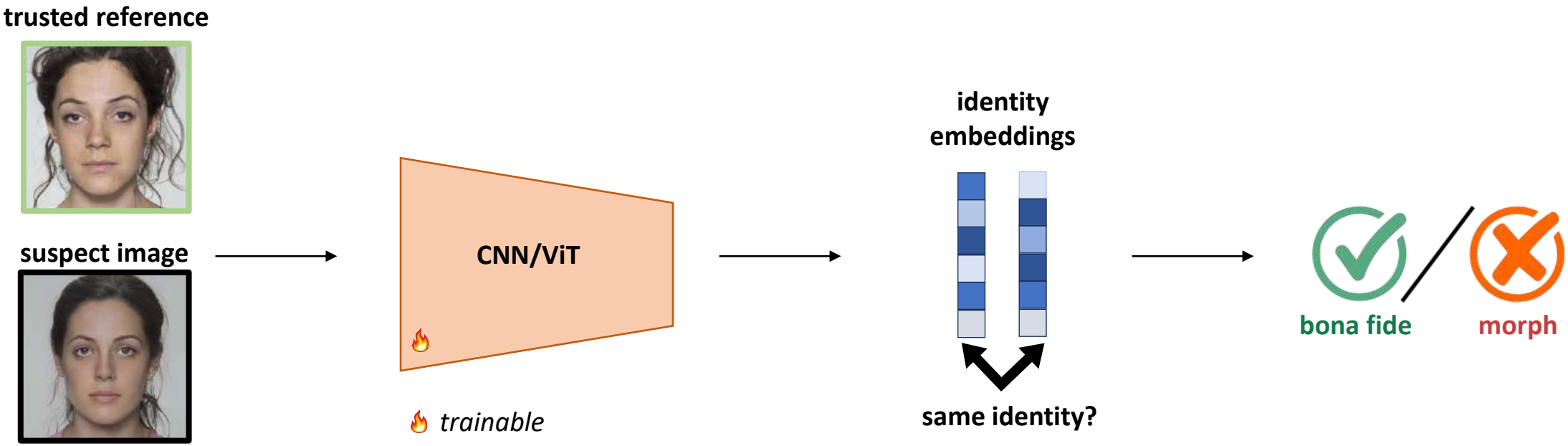
+ Less dependent on visible morph artifacts

– Sensitive to legitimate variation of one person's identity

Scherhag et al., Deep Face Representations for Differential MAD, IEEE TIFS, 2020

D-MAD: Pairwise Local Comparison

- **Compare multiple feature levels or local facial regions**. Local comparison may expose inconsistencies that a global embedding averages away.
- Unlike differential identity embedding methods which define the comparison first and learn a classifier afterwards, **pairwise models learn the comparison itself**.
- Modern methods increasingly replace conventional FR backbones with richer pretrained/foundation representations.



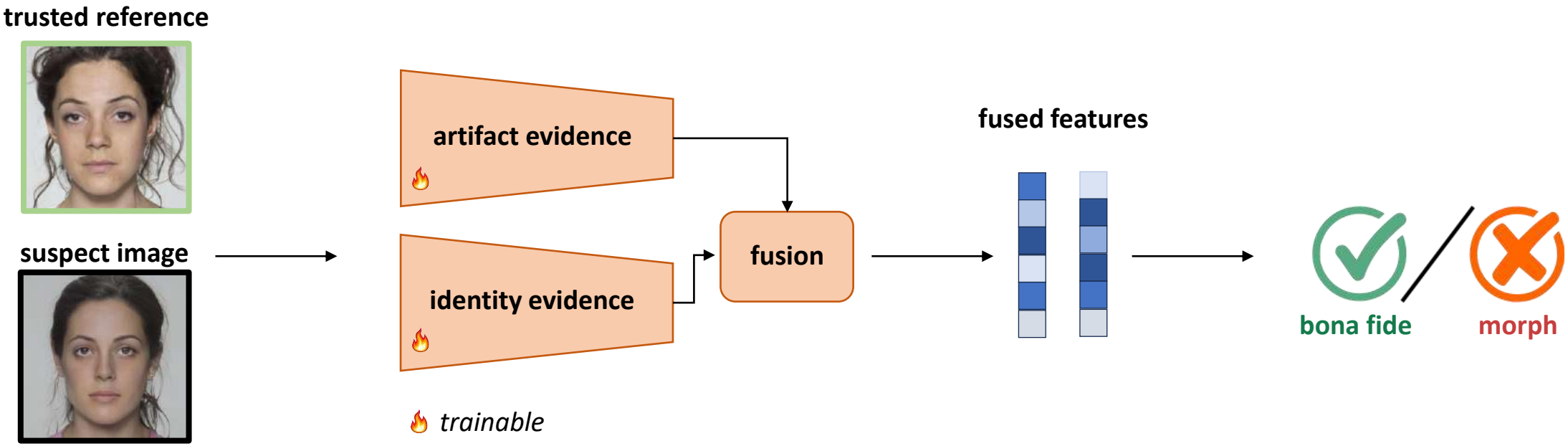
+ Can model complex and localized cross-image relationships

– Prone to shortcut learning (acquisition settings, pose, quality)

Gonzalez-Soler et al., DifFoundMAD: Foundation Models meet Differential MAD, IJCB, 2026

D-MAD: Fusion of Identity and Manipulation Evidence

- **Identity evidence**: how the claimed identity differs between document and live capture.
- **Artifact evidence**: S-MAD-like manipulation evidence present in the document relative to the trusted image.
- **Fusion** learns how to weight and combine identity and artifact evidence.



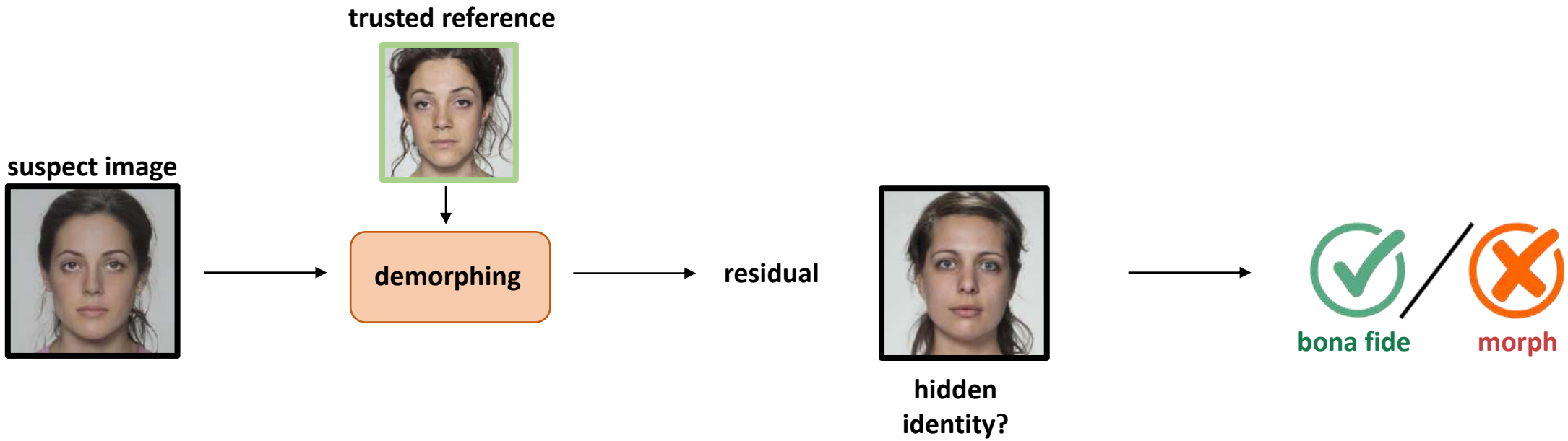
+ Complementary evidence can improve accuracy and robustness

– Conflicting cues can make fusion difficult

Borghi et al., A double siamese framework for differential morphing attack detection, Sensors, 2021

D-MAD: Demorphing

- Use the trusted reference to **remove that person's contribution from the suspect image**
- If the suspect image is morphed, demorphing can expose residual evidence of the other contributor.
- Learn whether the residual is consistent with same-identity variation or evidence of a second identity.



+ Provides explicit evidence of a hidden second identity

– Reconstruction errors may look like identity evidence

Ferrara et al., Face Demorphing, IEEE TIFS, 2018

D-MAD: Stronger Evidence, Different Sources of Ambiguity

1 Reference trust

A manipulated, injected, or non-live reference invalidates the premise.

2 Face appearance change

Large time gaps, facial hair, cosmetics, and surgery can raise false alarms.

3 Capture mismatch

Pose, illumination, blur, compression, and sensor differences distort embeddings.

4 Natural similarity

Twins, siblings, lookalikes, and carefully selected accomplices create identity ambiguity.

5 Local inconsistency ambiguity

Regional differences can arise from expression or capture rather than morphing.

**NIST FATE MORPH 2026 reports ageing-related false detections for some D-MAD algorithms.*

D-MAD: Demo

Double Siamese Detector

- Two complementary branches: artifact branch and identity branch
- **Artifact branch:**
  - Inception-ResNet V1 as a backbone
- **Identity branch:**
  - ArcFace face-recognition encoder for embedding extraction
  - Embeddings are subtracted
- An MLP combines artifact and identity evidence
  Supervised training on document/live pairs, learning complementary artifact and identity-discrepancy cues
- *Input*: document/probe image and a live/reference image
- *Output*: continuous D-MAD morph score

What we will test

1 — Detection
Bona fide document + trusted live image → morph document + trusted live image

2 — Robustness
Keep the document fixed → degrade the trusted live image → does the D-MAD score remain stable?

Borghi et al., A double siamese framework for differential morphing attack detection, Sensors, 2021

V-MAD: Moving From One Trusted Image to a Trusted Sequence

- Compare the document with **multiple live frames**, not only one
- **Each frame provides a different view** under varying pose, illumination, blur, face expression etc.
- Multiple observations can **reduce sensitivity to one poor-quality probe frame**

+

The first V-MAD study shows that even simple averaging of frame evidence improves over single-frame D-MAD.

Borghi et al., V-MAD: Video-based morphing attack detection in operational scenarios, IJCB, 2024

Session 3: Key Takeaway

1

Evidence determines the task

S-MAD, D-MAD, and V-MAD exploit fundamentally different evidence and should be designed accordingly.

2

Generalization remains challenging

Strong performance on known data does not guarantee robustness to unseen generators, datasets, post-processing, or subject populations.

3

Interpretability is not automatic

Heatmaps and natural-language explanations may be plausible without being well-grounded or faithful to the decision.

Evaluation of MADs

Evaluation: What Are We Measuring?

- **Detector accuracy:** *Can the detector separate bona fide and morphed samples?*
- **Generalization:** *Does performance hold for unseen attacks, datasets, and acquisition conditions?*
- **Operational usefulness:** *What false-alarm rate can the application tolerate?*

A single accuracy number is not enough to characterize a MAD system.

ISO/IEC 20059:2025. Methodologies to evaluate resistance of biometric systems to morphing attacks

Evaluation: Core MAD Metrics

MACER (**M**orphing **A**ttack **C**lassification **E**rror **R**ate):
Fraction of morphing attacks incorrectly classified as bona fide.

$$\text{MACER} = \frac{\text{morphs classified as bona fide}}{\text{all morphs}}$$

BSCER (**B**ona Fide **S**ample **C**lassification **E**rror **R**ate):
Fraction of bona fide samples incorrectly classified as morphs.

$$\text{BSCER} = \frac{\text{false morph alarms}}{\text{all bona fide samples}}$$

EER (**E**qual **E**rror **R**ate):
Error at **just one** operating point where MACER = BSCER.

ISO/IEC 20059:2025. Methodologies to evaluate resistance of biometric systems to morphing attacks

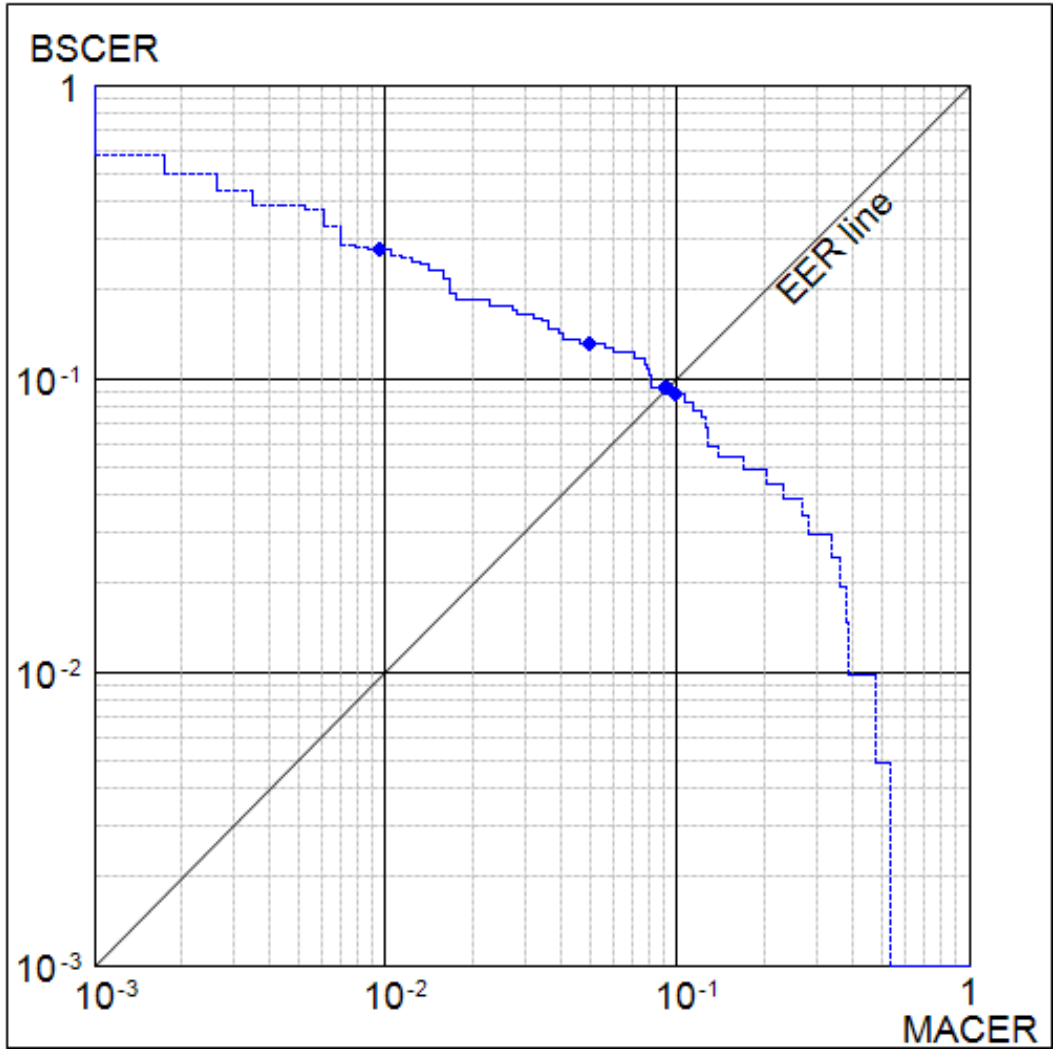
Evaluation: EER Is Not Enough

A **DET curve** shows the full threshold trade-off:

- x-axis: MACER
- y-axis: BSCER

EER is only a balanced reference point.
Real MAD deployments usually operate at thresholds chosen to constrain the more costly error.

Few common MAD operating points:
BSCER @ MACER = 10%, 5%, 1% · MACER @ BSCER = 0.3%



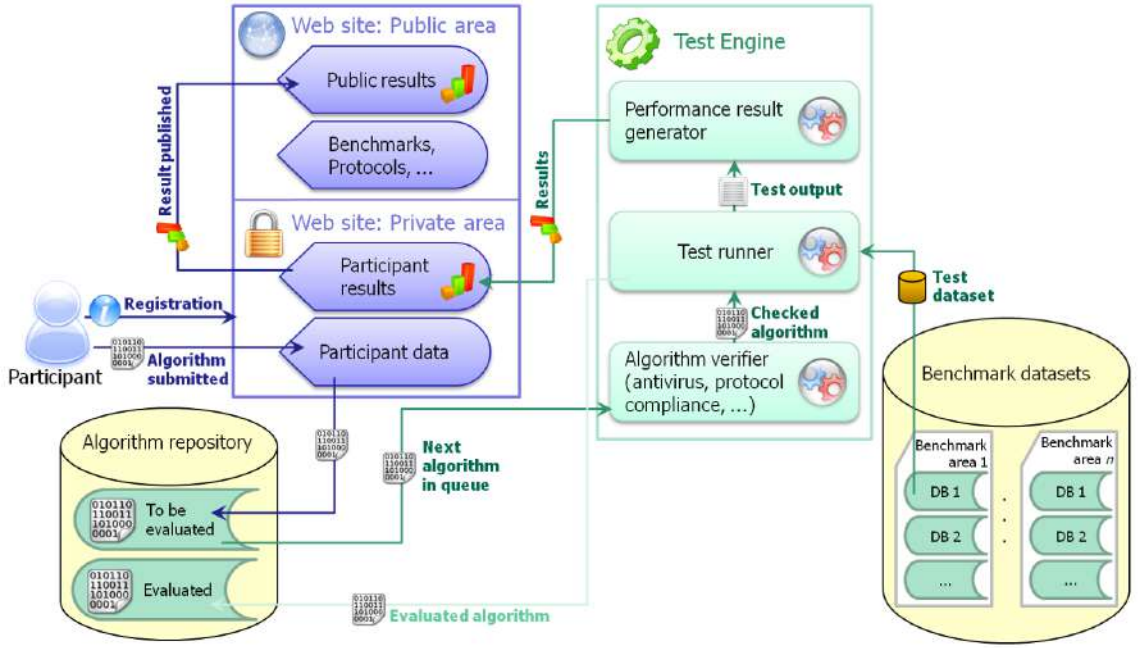
ISO/IEC 20059:2025. Methodologies to evaluate resistance of biometric systems to morphing attacks

Independent Evaluation: FVC-onGoing / BOEP (Bologna Online Evaluation Platform)

Goal: provide a **trusted** and **independent platform** for evaluating the performance of MAD approaches.

How it works:

- Participants can submit their algorithms in different formats (Win32 console application executables, Python scripts, Linux dynamically-linked libraries compliant to NIST FATE MORPH specifications).
- Evaluations are performed on sequestered datasets within a controlled execution environment.
- Performance is reported using common evaluation protocols and standardized metrics.
- Participants can decide to publish their results on a public leaderboard.



Independent Evaluation: FVC-onGoing / BOEP (Bologna Online Evaluation Platform)

Two benchmark MAD areas:

Single-image Morph Attack Detection

This benchmark area contains face morphing detection benchmarks. Morphing detection consists in analyzing an ISO compliant face image to determine whether it is the result of a morphing process (mixing faces of two subjects) or not. Algorithms submitted to these benchmarks are required to analyze an image and produce a score representing the probability of the image to be morphed. Read more...

Differential Morph Attack Detection

This benchmark area contains face morphing detection benchmarks. Morphing detection consists in analyzing an ISO compliant face image to determine whether it is the result of a morphing process (mixing faces of two subjects) or not. Algorithms submitted to these benchmarks are required to compare a bona fide (not morphed) image to a suspected image and produce a score representing the probability of the suspected image to be morphed. Read more...

Independent Evaluation: FVC-onGoing / BOEP (Bologna Online Evaluation Platform)

S-MAD Benchmarks:

Year	Benchmarks	Format	Bona Fide Attempts	Morphing Attempts
2016	SMAD-BIOLAB-1.0	Digital	88	80
2018	SMAD-MORPHDB_D-1.0	Digital	130	100
	SMAD-MORPHDB_P&S-1.0	Printed & Scanned	130	100
2020	SMAD-SOTAMD_D-1.0	Digital	300	2045
	SMAD-SOTAMD_P&S-1.0	Printed & Scanned	1096	3703
	SMAD-SOTAMD_PM_D-1.0	Digital	300	470
	SMAD-SOTAMD_UC_P&S-1.0	Printed & Scanned	200	380
2023	SMAD-IMARS-HQ_FULL-1.0	Digital	300	25200
	SMAD-IMARS-HQ_SMALL-1.0	Digital	300	5040
	SMAD-IMARS-HQ_HARD-1.0	Digital	300	6522

Independent Evaluation: FVC-onGoing / BOEP (Bologna Online Evaluation Platform)

D-MAD Benchmarks:

Year	Benchmarks	Format	Bona Fide Attempts	Morphing Attempts
2016	DMAD-BIOLAB-1.0	Digital	526	160
2018	DMAD-MORPHDB_D-1.0	Digital	756	396
	DMAD-MORPHDB_P&S-1.0	Printed & Scanned	756	396
2020	DMAD-SOTAMD_D-1.0	Digital	3000	30550
	DMAD-SOTAMD_P&S-1.0	Printed & Scanned	10960	55530
	DMAD-SOTAMD_PM_D-1.0	Digital	3000	7050
	DMAD-SOTAMD_UC_P&S-1.0	Printed & Scanned	2000	5700
2023	DMAD-IMARS-HQ_FULL-1.0	Digital	3000	504000
	DMAD-IMARS-HQ_SMALL-1.0	Digital	3000	50400
	DMAD-IMARS-HQ_HARD-1.0	Digital	3000	130440

Independent Evaluation: FVC-onGoing / BOEP (Bologna Online Evaluation Platform)

Results: example

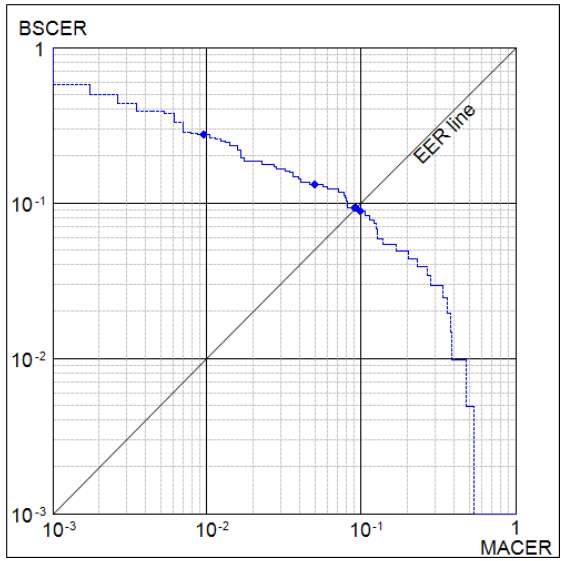
$BSCER_x$ is the lowest $BSCER$ for $MACER \leq \frac{1}{x}\%$

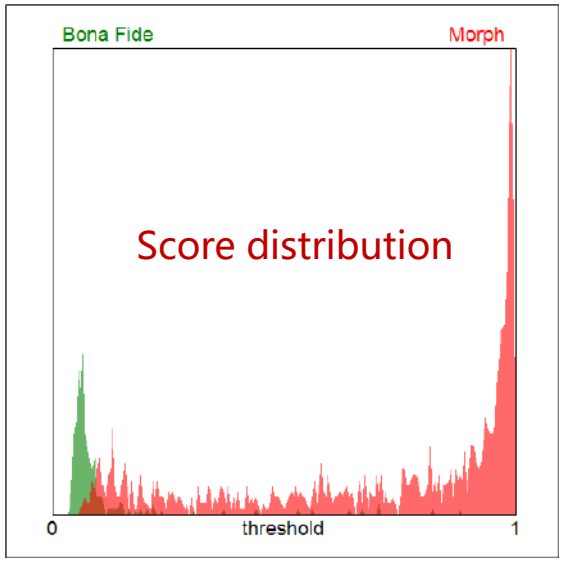
Main performance indicators

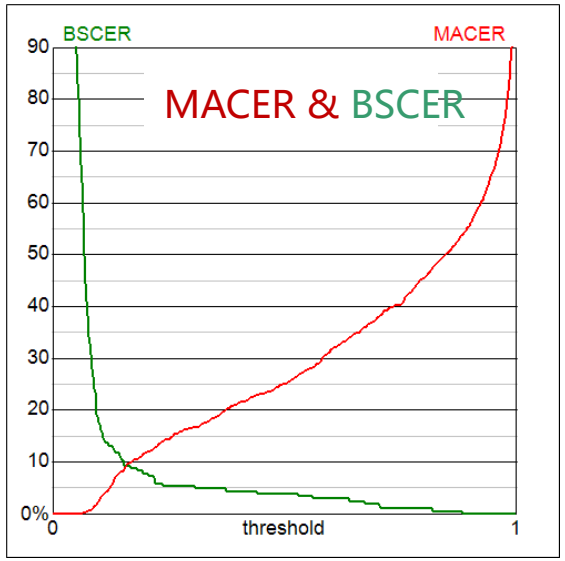
Accuracy indicators			
EER	$BSCER_{10}$	$BSCER_{20}$	$BSCER_{100}$
9.23% (9.19% - 9.27%)	8.78%	13.17%	27.32%

Detection failures	
REJ_{NBFRA}	REJ_{NMRA}
0%	0%

Detailed Graphs







Independent Evaluation: NIST FATE MORPH

- Independent evaluation of **S-MAD** and **D-MAD** technologies
- Uses sequestered morph datasets spanning **different morph-generation techniques** and difficulty levels
- Strong emphasis on **operational operating points**, rather than only EER
- Includes **analysis of image quality, ageing, print/scan and attack characteristics**
- Results and operational guidance are publicly reported

Session 4: Evaluation Key Takeaway

1

Metrics need context

EER is useful for comparison, but deployment requires application-specific operating points.

2

Generalization must be demonstrated

Unseen generators, datasets, acquisition conditions and populations matter more than closed-set accuracy.

3

Independent evaluation matters

Sequestered platforms such as FVC-onGoing and NIST reduce benchmark overfitting and make comparisons more credible.