NISTIR 8311

Ongoing Face Recognition Vendor Test (FRVT)

Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms

Mei Ngan
Patrick Grother
Kayee Hanaoka

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8311

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 8311

Ongoing Face Recognition Vendor Test (FRVT)

Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms

Mei Ngan
Patrick Grother
Kayee Hanaoka
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8311

July 2020

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

National Institute of Standards and Technology Interagency or Internal Report 8311
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8311, 58 pages (July 2020)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8311

Executive Summary

OVERVIEW

This is the first of a series of reports on the performance of face recognition algorithms on faces occluded by protective face masks [2] commonly worn to reduce inhalation of viruses or other contaminants. This study is being run under the Ongoing Face Recognition Vendor Test (FRVT) executed by the National Institute of Standards and Technology (NIST). This report documents accuracy of algorithms to recognize persons wearing face masks. The results in this report apply to algorithms provided to NIST before the COVID-19 pandemic, which were developed without expectation that NIST would execute them on masked face images. NIST had informed the FRVT developer community of our intent to run existing algorithms on masked images prior to the outset of this study and invited submission of mask-enabled algorithms for the next phase of this work. This report is intended to support end-users to understand how a pre-pandemic algorithm might be affected by the arrival of a substantial number of subjects wearing face masks. The next report will document accuracy values for more recent algorithms, some developed with capabilities for recognition of masked faces. The algorithms tested were one-to-one algorithms submitted to the FRVT 1:1 Verification track. Future iterations of this document will also report accuracy of one-to-many algorithms.

MOTIVATION

Traditionally, face recognition systems (in cooperative settings) are presented with mostly non-occluded faces, which include primary facial features such as the eyes, nose, and mouth. However, there are a number of circumstances in which faces are occluded by masks such as in pandemics, medical settings, excessive pollution, or laboratories. Inspired by the COVID-19 pandemic response, the widespread requirement that people wear protective face masks in public places has driven a need to understand how cooperative face recognition technology deals with occluded faces, often with just the periocular area and above visible. An increasing number of research publications have surfaced on the topic of face recognition on people wearing masks along with face-masked research datasets [7]. Several commercial providers have recently developed ”face mask capable” face recognition systems which were not tested in this report. Results for such face mask capable or post-COVID algorithms will be published in the next report of this face mask evaluation series. This report documents results for pre-COVID algorithms developed primarily for non-covered faces, comparing an unmasked portrait quality enrollment image to a synthetically-masked webcam probe image. To date, we are not aware of any large-scale, independent, and publicly reported evaluation on the effects of face mask occlusion on face recognition.

WHAT WE DID

The NIST Information Technology Laboratory (ITL) quantified the accuracy of pre-COVID face recognition algorithms on faces occluded by masks applied digitally to a large set of photos that has been used in an FRVT verification benchmark since 2018. These algorithms were submitted to FRVT 1:1 prior to the COVID-19 pandemic and were developed without expectation that NIST would execute them on masked face images. Using the original unmasked images to form a baseline for accuracy, we measured the impact of occlusion by digitally applying a mask to the face and varying mask shape, mask color, and nose coverage.

We used these algorithms with two large datasets of photographs collected in U.S. governmental applications that are currently in operation: unmasked **application photographs** from a global population of applicants for immigration benefits and digitally-masked **border crossing photographs** of travelers entering the United States. Both datasets were collected for authorized travel or immigration processes. The application photos (used as reference images) have good compliance with image capture standards. The digitally-masked border crossing photos (used as probe images) are not in good compliance with image capture standards given constraints on capture duration and environment. The application photos were left unmasked, and synthetic masks were applied to the border crossing photos. This mimics an operational scenario where a person wearing a mask attempts to authenticate against a prior visa or passport photo. Together these datasets allowed us to process a total of 6.2 million images of 1 million people through 89 algorithms.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

WHAT WE DID (CONTINUED)

Our use of software to apply masks to face images has the following advantages: it allows very rapid characterization of the effect of masks on face recognition; it allows controlled exploration of factors such as mask size, shape, and color; it affords repeatability, which is key to the fair comparison of algorithms; it scales to very large datasets - in our study, some 6.2 million photographs - which allows fine-grained characterization of false positive rates in addition to false negative rates. Inversely, our use of digital masks presents a number of limitations: our digital masks are tailored to faces whereas realistically, mass-produced real masks may fit differently on different people; our use of uniformly-colored masks does not capture the impact of mask texture or pattern on face recognition; we were not able to pursue an exhaustive simulation of the endless variations in color, design, shape, texture, bands, and ways masks can be worn; our study does not capture any camera-mask interactions which may cause over or underexposure of the periocular region or face detection issues. Please see the *Limitations* section of this executive summary for a more detailed discussion on the limitations of this study.

WHAT WE FOUND

The following results apply to algorithms provided to NIST before the COVID-19 pandemic, which were developed without expectation that NIST would execute them on masked face images. The study has certain limitations, and these are discussed in the next section.

▷ **False rejection performance**: All algorithms give increased false non-match rates when the probes are masked. Using border crossing images, without masks, the most accurate algorithms will fail to authenticate about 0.3% of persons while falsely accepting no more than 1 in 100000 impostors (i.e. FNMR= 0.003 at FMR= 0.00001). With the highest coverage mask we tested and the most accurate algorithms, this failure rate rises to about 5% (FNMR = 0.05). This is noteworthy given that around 70% of the face area is occluded by the mask. However, many algorithms are much less tolerant: some algorithms that are quite competitive with unmasked faces (FNMR < 0.01) fail to authenticate between 20% and 50% of images (FNMR $\rightarrow 0.5$). *See Table 3 and Figure 3*

In cooperative access control applications, false negatives can traditionally be remedied by users making second attempts. This is effective when users correct pose, expression, or illumination aspects of their presentation. With masked faces, however, a second attempt may not be effective if the failure is a systematic property of the algorithm.

▷ **False acceptance performance**: As most systems are configured with a fixed threshold, it is necessary to report both false negative and false positive rates for each group at that threshold. In most cases false match rates are reduced by masks. The effect is generally modest with reductions in FMR usually being smaller than a factor of two. This property is valuable in that masks do not impart adverse false match security consequences for verification. *See Figure 27*

▷ **Coverage of the masks**: Unsurprisingly masks that occlude more of the face give larger false non-match rates. We surveyed over the extent to which the mask covers the nose, from not at all ("low") to typical ("medium") to near the eyes ("high"). We baselined those with unmasked faces with the result that FNMR increases by factors of around 10, 25, and 36 respectively for the median algorithm. However, as noted, algorithms vary considerably in the their tolerance of coverage. Readers should consult tabulated values for specific algorithms. *See Table 3 and Figure 3*

We included the "low" option not because it is a common position for a mask but as an option for authentication applications where it would be tenable to ask the user to pull the mask down to just below the nose for the duration of the authentication attempt.

▷ **Shape of the masks**: The shape of the masks matters. Full-face-width masks generally cover more of the face than rounder N95 type masks. Results show that wide-width masks generally give false negative rates about a factor of two higher than do rounder type masks. *See Figure 14*

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

WHAT WE FOUND (CONTINUED)

- **Color of the masks**: We considered light-blue and black masks. Most algorithms have higher error rates in black masks than light-blue masks. The reason for observed accuracy differences between mask color is unknown but is a point for consideration by impacted developers. Mask color also affects the rate at which some algorithms fail to produce a template from an image. *See Table 5*
- **Failure to detect and template:** The false negative rates in this report include the effects of both face detection and localization errors, and low-similarity matching errors. We separately include tables detailing how often an algorithm does not make a template from an input image. This can occur because the algorithm doesn't detect a face, or electively chooses not to extract features from it. While many algorithms give low failure-to-template rates, some give high values ranging up to 100%. Inversely, the successful creation of a template does not guarantee proper facial localization (e.g. algorithms may incorrectly detect something that's not a face). Such localization failures will not be captured as a failure to detect and template event but will impact accuracy rates nonetheless. *See Table 5*

LIMITATIONS

As a simulation, this study likely doesn't fully capture the effects of masks on face recognition. Particularly the following points should be weighed by readers in the near term. Some of these will be addressed in subsequent work at NIST.

- **Evaluate "mask-enabled" algorithms:** The algorithms used so far were submitted to the FRVT by corporate research and development laboratories and a few universities in 2019 and early 2020. Several of the algorithms were submitted to NIST as recently as March 2020, but because the algorithms were developed without expectation that NIST would run them on faces occluded by masks, we consider all algorithms evaluated here as "pre-pandemic".
- **Apply masks to both photos:** We masked only the probe image. We did not mask the reference photo. This situation represents authentication against an unmasked photo drawn from a pre-pandemic credential (e.g. passport) or database. While in some applications masks could appear on both enrollment and recognition images, we anticipate "mask-enabled" algorithms will need to extract and compare features from all combinations of masked and unmasked photos.
- **Train algorithms:** As with all NIST evaluations, we regard the software as a black box whose parameters (models) remain fixed for the entirety of its use without learning from the test data. We do not train or fine-tune algorithms.
- **Evaluate one-to-many algorithms:** We have only run one-to-one verification algorithms with masks. This elicits data on the effect of masks on the underlying feature extraction and discrimination of algorithms and can therefore be be expected to give first-order indications of the effect on one-to-many identification algorithms.
- **Consider the effect of eye occlusion:** We did not address the effect of eye-glasses or eye-protection. While our dataset includes examples of people wearing glasses, we didn't collect such data nor simulate it with digital addition.
- **Test with images of real masks:** Given time and resource constraints, we didn't collect photos of subjects wearing masks. The possible downsides of this are several. First, our digital masks are tailored to faces; while a few don't fit realistically, mass-produced real masks may not fit all actual persons correctly either. Second, because many cameras run with exposure-control, it is possible that a dark mask will cause less light to be reflecting and the camera to increase gain on the sensor causing overexposure of the periocular region. Likewise a white mask could lead to underexposure problems. Third, it is possible that some cameras that include a face detector, may fail to focus or acquire a masked face correctly.

LIMITATIONS (CONTINUED)

- **Use textured masks:** All masks synthesized by NIST in this study have a uniform color. The consequences of this are that we do not capture the the increasing diversity of masks worn recently, including those with corporate logos, text, patterns, or those advertised to thwart face recognition. The possibility exists for patterned masks to induce higher facial localization errors, which is not captured in our current study. We received a suggestion that such information may serve as a soft biometric, in that a subject that always wears the same textured mask will be more identifiable. We don't intend to encourage algorithm development along this line, because as mass-produced high-efficacy masks become more common, mask diversity may actually drop.
- **Study demographic effects on masked images:** This report estimates overall performance of existing algorithms on recognition of faces occluded by masks. We deferred tabulating accuracy for different demographic groups until more capable mask-enabled algorithms have been submitted to FRVT.
- **Evaluate algorithms on non-cooperative, unconstrained imagery:** This report documents results for matching masked webcam images to unmasked portrait-style photos. While the properties of the two sets of images differ, subjects are operating in cooperative mode and are for the most part, looking at the camera.
- **Consider effects of human examination:** This report does not consider the various ways humans are involved in face recognition systems. For example, analysts can correct face detection or localization errors induced by masks, prior to automated recognition. Likewise, humans are often tasked with adjudication of images following a rejection or other exception from an automated system. Analysis of human capability and role is pertinent to those operations, but is beyond the scope of this study.

IMPLICATIONS AND FUTURE WORK

Know Your Algorithm: Operational implementations usually employ a single face recognition algorithm. Given algorithm-specific variation, it is incumbent upon the system owner to know their algorithm. While publicly available test data from NIST and elsewhere can inform owners, it will usually be informative to specifically measure accuracy of the operational algorithm on the operational image data collected with actual masks.

NIST plans on releasing a series of reports, iteratively assessing different aspects and use cases of face masking on recognition performance. In the near term, we anticipate the next report in this series to evaluate the performance of "mask-enabled" algorithms submitted to FRVT.

ACKNOWLEDGMENTS

This work was conducted in collaboration with the Department of Homeland Security's Science & Technology Directorate (S&T), Office of Biometric Identity Management (OBIM), and Customs and Border Protection (CBP). Additionally, the authors are grateful to staff in the NIST Biometrics Resesarch Laboratory for infrastructure supporting rapid evaluation of algorithms.

DISCLAIMER

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

INSTITUTIONAL REVIEW BOARD

The National Institute of Standards and Technology's Research Protections Office reviewed the protocol for this project and determined it is not human subjects research as defined in Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule for the Protection of Human Subjects (45 CFR 46, Subpart A).

Contents

List of Tables

List of Figures

1 Face Mask Effects

1.1 Status

NIST has conducted the first out of a series of tests aimed at quantifying face recognition accuracy for people wearing masks. Our initial approach has been to apply masks to faces digitally (i.e., using software to apply a synthetic mask). This allowed us to leverage large datasets that we already have. This initial report documents results for 1:1 verification algorithms. We tested algorithms that were already submitted to FRVT 1:1 prior to mid-March 2020. The results in this report apply to algorithms provided to NIST **before the COVID-19 pandemic** and for which developers had no expectation that NIST would execute them on masked face images. This report is intended to support end-users to understand how a pre-pandemic algorithm might be affected by the arrival of substantial number of subjects wearing face masks. The next report will document accuracy values for more recent algorithms, some developed with capabilities for recognition of masked faces. These initial results reflect the case when only the probe image is masked and the reference photo is left as-is. Future reports will consider the effect of masking both enrollment and verification images. This report quantifies the effect of masks on both false negative and false positives match rates.

The FRVT evaluation is an ongoing test that remains open to new participation. Comments and suggestions should be directed to frvt@nist.gov.

1.2 Introduction

The majority of face recognition systems have been deployed in applications where subjects make cooperative presentations to a camera, for example as part of an application for a benefit or ID credential, or as during access control. With very few exceptions such images would not include face masks or other occlusions. However, with the SARS-CoV-2 pandemic, we can anticipate a demand to authenticate persons wearing masks, for example in immigration settings, without the need to the subjects to remove those masks. This presents a problem for face recognition, because regions of the face occluded by masks - the mouth and nose - include information useful for both recognition and, potentially, the detection stage that precedes it.

Previous work on face recognition of occluded faces has been directed at situations such as crime scenes where subjects were actively un-cooperative i.e. acting to evade face detection and recognition. Those applications are often characterized by image properties (low resolution, video compression, uncontrolled head orientation) that are known [4] to degrade recognition accuracy.

2 Image Datasets

2.1 Application Images

The images are collected in an attended interview setting using dedicated capture equipment and lighting. The images, at size 300x300 pixels. The images are all high-quality frontal portraits collected in immigration offices and with a white background. As such, potential quality related drivers of high false match rates (such as blur) can be expected to be absent. The images are encoded as ISO/IEC 10918 i.e. JPEG. Over a random sample of 1000 images, the images have compressed file sizes (mean: 42KB, median: 58KB, 25-th percentile: 15KB, and 75-th percentile: 66KB). The implied bit-rates are mostly benign and superior to many e-Passports. When these images are provided as input into the algorithm, they are labeled with the type "ISO". This report used 1 019 232 application images.

Figure 1: Examples of images with properties similar to the enrollment application photos used in this study. The subjects in the photos are all NIST employees.

2.2 Webcam Images

These images are taken with a camera oriented by an attendant toward a cooperating subject. This is done under time constraints, so there are roll, pitch, and yaw angle variations. Also, background illumination is sometimes bright, so the face is under exposed. Sometimes, there is perspective distortion due to close range images. The images are generally in poor conformance with the ISO/IEC 19794-5 Full Frontal image type. The images have mean interocular distance of 38 pixels. The images are all live capture. When these images are provided as input into the algorithm, they are labeled with the type "WILD". Examples of such images are included in Figure 2 and Figure 4 in NIST Interagency Report 8271. Results for verification of these images (unmasked) appear in FRVT Part 1 - Verification both compared against images of the same type, and with those described in section 2.1. This report used 5 225 633 border webcam images.

2.3 Synthetically Masked Images

In this test, synthetically-generated masks were overlaid on top of all probe images, which in this case, were webcam images described in Section 2.2. The Dlib [5] C++ toolkit version 19.19 was used to detect and establish key facial points on the face, and with the facial points, solid masks of different shape, height, and color were drawn on the face. The exact Dlib facial points and details used to generate the masks are documented in Appendix A. In the event that Dlib was unable to detect a face in the image, eye coordinates were used to generate a mask leveraging standardized token frontal geometry [1].

Examples of unmasked enrollment images and synthetically-masked probe images are presented in Figures 1 and 2, respectively.

1 ORIGINAL IMAGE

2 WIDE, HIGH COVERAGE

3 WIDE, MEDIUM COVERAGE

4 WIDE, LOW COVERAGE

5 ROUND, HIGH COVERAGE

6 AS ROW 3 IN BLACK

Figure 2: Examples of synthetically-generated face masks used in this study. The original images are from the NIST MEDS-II Dataset [3]. They were collected in operational settings using the same camera and procedure as is used for the border images that form the mainstay of the experiments in this report.

3 Metrics

3.1 Matching accuracy

Given a vector of N genuine scores, u, the false non-match rate (FNMR) is computed as the proportion below some threshold, T:

$$\text{FNMR}(T) = 1 - \frac{1}{N}\sum_{i=1}^{N} H(u_i - T) \tag{1}$$

where $H(x)$ is the unit step function, and $H(0)$ taken to be 1.

Similarly, given a vector of N impostor scores, v, the false match rate (FMR) is computed as the proportion above T:

$$\text{FMR}(T) = \frac{1}{N}\sum_{i=1}^{N} H(v_i - T) \tag{2}$$

The threshold, T, can take on any value. We typically generate a set of thresholds from quantiles of the observed impostor scores, v, as follows. Given some interesting false match rate range, $[\text{FMR}_L, \text{FMR}_U]$, we form a vector of K thresholds corresponding to FMR measurements evenly spaced on a logarithmic scale

$$T_k = Q_v(1 - \text{FMR}_k) \tag{3}$$

where Q is the quantile function, and FMR_k comes from

$$\log_{10}\text{FMR}_k = \log_{10}\text{FMR}_L + \frac{k}{K}\,\left[\log_{10}\text{FMR}_U - \log_{10}\text{FMR}_L\right] \tag{4}$$

Error tradeoff characteristics are plots of FNMR(T) vs. FMR(T). These are plotted with $\text{FMR}_U \rightarrow 1$ and FMR_L as low as is sustained by the number of impostor comparisons, N. This is somewhat higher than the "rule of three" limit $3/N$ because samples are not independent, due to re-use of images.

3.2 Failure to Enroll

Failure to enroll (FTE) is the proportion of failed template generation attempts. Failures can occur because the software throws an exception, or because the software electively refuses to process the input image. This would typically occur if a face is not detected. FTE is measured as the number of function calls that give EITHER a non-zero error code OR that give a "small" template. This is defined as one whose size is less than 60 bytes. This second rule is needed because some algorithms incorrectly fail to return a non-zero error code when template generation fails yet do return a valid default data structure.

The effects of FTE are included in the accuracy results of this report by regarding any template comparison involving a failed template to produce a low similarity score. Thus higher FTE results in higher FNMR and lower FMR.

4 Algorithms

The FRVT activity is open to participation worldwide, and the test will evaluate submissions on an ongoing basis. There is no charge to participate. The process and format of algorithm submissions to NIST are described in the FRVT 1:1 Verification Application Programming Interface (API) [6] document. Participants provide their submissions in the form of libraries compiled on a specific Linux kernel, which are linked against NIST's test harness to produce executables. NIST provides a validation package to participants to ensure that NIST's execution of submitted libraries produces the expected output on NIST's test machines.

This report documents the results of algorithms submitted to FRVT 1:1 for testing from April 2019 to March 2020, without specific claims to being able to recognize people wearing face masks. Table 2 lists the algorithms that were tested. Note that algorithms that expired prior to June 2020 were not included in this report.

	Developer	Algorithm	Submission Date
1	3Divi	3divi-004	2019-07-22
2	ADVANCE.AI	advance-002	2019-12-19
3	ASUSTek Computer Inc	asusaics-000	2019-10-24
4	Ability Enterprise Co. Ltd - Andro Video	androvideo-000	2020-02-03
5	Acer Incorporated	acer-000	2020-01-08
6	Ai First	aifirst-001	2019-11-21
7	AiUnion Technology Co Ltd	aiunionface-000	2019-10-22
8	AlphaSSTG	alphaface-002	2020-02-20
9	Anke Investments	anke-005	2019-11-21
10	Antheus Technologia Ltda	antheus-000	2019-12-05
11	Aware	aware-005	2020-02-27
12	Awidit Systems	awiros-001	2019-09-23
13	Beijing Alleyes Technology Co Ltd	alleyes-000	2020-03-09
14	BioID Technologies SA	bioidtechswiss-000	2019-11-15
15	CSA IntelliCloud Technology	intellicloudai-001	2019-08-13
16	CTBC Bank Co Ltd	ctbcbank-000	2019-06-28
17	Camvi Technologies	camvitech-004	2019-07-12
18	Canon Information Technology (Beijing) Co Ltd	cib-000	2019-12-11
19	China University of Petroleum	upc-001	2019-06-05
20	Chinese Univeristy of Hong Kong	cuhkee-001	2020-03-18
21	Chosun University	chosun-000	2020-02-12
22	Chunghwa Telecom Co. Ltd	chtface-002	2019-12-07
23	Cyberlink Corp	cyberlink-004	2020-02-27
24	DSK	dsk-000	2019-06-28
25	Dahua Technology Co Ltd	dahua-004	2019-12-18
26	Deepglint	deepglint-002	2019-11-15
27	DiDi ChuXing Technology Co	didiglobalface-001	2019-10-23
28	Expasoft LLC	expasoft-000	2020-01-06
29	FaceSoft Ltd	facesoft-000	2019-07-10
30	Fujitsu Research and Development Center Co Ltd	fujitsulab-000	2020-02-04
31	Glory Ltd	glory-002	2019-11-12
32	Gorilla Technology	gorilla-005	2020-03-11
33	Guangzhou Pixel Solutions Co Ltd	pixelall-003	2019-10-15
34	ITMO University	itmo-007	2020-01-06
35	Idemia	Idemia-005	2019-10-11
36	Imagus Technology Pty Ltd	imagus-001	2019-10-22
37	Imperial College London	imperial-002	2019-08-28
38	Incode Technologies Inc	incode-006	2020-02-20
39	Innovative Technology Ltd	innovativetechnologyltd-002	2020-02-26
40	Innovatrics	innovatrics-006	2019-08-13
41	Institute of Information Technologies	iitvision-002	2019-12-04
42	Intel Research Group	intelresearch-001	2020-01-14
43	Intellivision	intellivision-002	2019-08-23
44	Kakao Enterprise	kakao-003	2020-02-26
45	Kedacom International Pte	kedacom-000	2019-06-03
46	Kneron Inc	kenron-005	2020-02-21
47	Lomonosov Moscow State University	intsysmsu-002	2020-03-12
48	Lookman Electroplast Industries	lookman-004	2019-06-03
49	Luxand Inc	luxand-000	2019-11-07
50	MVision	mvision-001	2019-11-12
51	Momentum Digital Co Ltd	sertis-000	2019-10-07
52	Moontime Smart Technology	mt-000	2019-06-03
53	N-Tech Lab	ntech-008	2020-01-06
54	Netbridge Technology Incoporation	netbridgetech-001	2020-01-08
55	Neurotechnology	neurotech-008	2020-01-08
56	Nodeflux	nodeflux-002	2019-08-13
57	NotionTag Technologies Private Limited	notiontag-000	2019-06-12

Table 1: List of algorithms included in this report.

	Developer	Algorithm	Submission Date
58	Panasonic R+D Center Singapore	psl-004	2020-03-03
59	Paravision (EverAI)	paravision-004	2019-12-11
60	Rank One Computing	rankone-008	2019-11-12
61	Remark Holdings	remarkai-001	2019-11-21
62	Rokid Corporation Ltd	rokid-000	2019-08-01
63	Samsung S1 Corp	s1-001	2019-12-06
64	Scanovate Ltd	scanovate-001	2019-11-12
65	Sensetime Group Ltd	sensetime-003	2019-06-04
66	Shanghai Jiao Tong University	sjtu-002	2020-02-12
67	Shanghai Ulucu Electronics Technology Co. Ltd	uluface-002	2019-07-10
68	Shanghai Universiy - Shanghai Film Academy	shu-002	2019-12-10
69	Shenzhen AiMall Tech Ltd	aimall-002	2020-03-12
70	Shenzhen Intellifusion Technologies Co Ltd	intellifusion-002	2020-03-18
71	Star Hybrid Limited	starhybrid-001	2019-06-19
72	Synology Inc	synology-000	2019-10-23
73	TUPU Technology Co Ltd	tuputech-000	2019-10-11
74	Taiwan AI Labs	ailabs-001	2019-12-18
75	Tech5 SA	tech5-004	2020-03-09
76	Tencent Deepsea Lab	deepsea-001	2019-06-03
77	Tevian	tevian-005	2019-09-21
78	Trueface.ai	trueface-000	2019-10-08
79	Universidade de Coimbra	visteam-000	2020-01-14
80	Via Technologies Inc	via-001	2020-01-08
81	Videmo Intelligente Videoanalyse	videmo-000	2019-12-19
82	Videonetics Technology Pvt Ltd	videonetics-002	2019-11-21
83	Vigilant Solutions	vigilant-007	2019-06-27
84	VisionLabs	visionlabs-008	2020-01-06
85	Vocord	vocord-008	2020-01-031
86	Winsense Co Ltd	winsense-001	2019-10-16
87	X-Laboratory	x-laboratory-001	2020-01-21
88	Xforward AI Technology Co LTD	xforwardai-000	2020-02-06
89	iQIYI Inc	iqface-000	2019-06-04

Table 2: List of algorithms included in this report.

5 Results

This section includes accuracy results for the 89 one-to-one verification algorithms listed in section 4. We do not include speed and computational resource requirements - they are given in Table 1 in the FRVT 1:1 report. The results, which span many pages, are comprised of:

- **FNMR**: Table 3 tabulates false non-match rates by color, shape and nose coverage. It includes also FNMR without any mask. FNMR values are stated at a fixed threshold calibrated to give FMR = 0.00001 on unmasked images.
- **DET**: Figure 3 shows detection error trade of characteristics spanning false match rates from $3\,10^{-7}$ to 1.
- **Mask vs. no mask**: The scatter plot in Figure 13 shows variation across all algorithms of FNMR without masks against FNMR with a common type of mask.
- **Mask shape**: The scatter plot in Figure 14 shows for all algorithms the increase in false negative results for wide masks vs. narrower round masks.
- **Nose coverage**: The scatter plot in Figure 15 shows for all algorithms the increase in false negative rates for masks that substantially cover the nose and those pulled beneath the nose.
- **FTE**: Table 5 gives empirical failure-to-template results by color, shape, and nose coverage. The table was produced using 10 000 images of each kind of mask.
- **FTE as contributor to FNMR**: The FNMR results include failure-to-template rates (FTE). Figure 16 shows the proportion of template generation failures.
- **FNMR vs. threshold**: Figure 17 shows explicit dependence of false non-match rate on threshold.
- **FMR vs. threshold**: Likewise Figure 27 shows explicit dependence of false match rate on threshold.

	Algorithm Name	NOT MASKED	MASKED COLOR = LIGHTBLUE						MASKED COLOR = BLACK		
			SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE		
	COVERAGE		LO	MED	HI	LO	MED	HI	LO	MED	HI
1	3divi-004	0.0130^{49}	0.4123^{55}	0.6760^{68}	-	-	-	-	-	-	-
2	acer-000	0.8432^{85}	0.9995^{68}	0.9999^{85}	-	-	-	-	-	-	-
3	advance-002	0.0328^{69}	-	0.2351^{24}	-	-	-	-	-	-	-
4	aifirst-001	0.0079^{29}	0.0778^{28}	0.2567^{27}	-	-	-	-	-	-	-
5	ailabs-001	0.0243^{62}	-	0.6792^{69}	-	-	-	-	-	-	-
6	aimall-002	0.0133^{51}	-	0.3919^{44}	-	-	-	-	-	-	-
7	aiunionface-000	0.0094^{34}	0.0917^{34}	0.2935^{35}	-	-	-	-	-	-	-
8	alleyes-000	0.0044^{7}	-	0.1038^{10}	-	0.0181^{8}	0.0542^{10}	0.1050^{10}	0.0262^{11}	0.1287^{13}	0.1991^{13}
9	alphaface-002	1.0000^{88}	1.0000^{70}	1.0000^{88}	-	-	-	-	-	-	-
10	androvideo-000	0.0333^{70}	0.3177^{51}	0.6498^{65}	-	-	-	-	-	-	-
11	anke-005	0.0062^{23}	0.0671^{21}	0.3207^{39}	-	-	-	-	-	-	-
12	antheus-000	0.7319^{84}	0.9994^{67}	0.9999^{84}	-	-	-	-	-	-	-
13	asusaics-000	0.0090^{33}	-	0.3616^{42}	-	-	-	-	-	-	-
14	aware-005	0.0308^{68}	0.4962^{57}	0.8876^{75}	-	-	-	-	-	-	-
15	awiros-001	0.1233^{76}	0.6823^{60}	0.8635^{74}	-	-	-	-	-	-	-
16	bioidtechswiss-000	0.0050^{10}	0.0308^{10}	0.1155^{12}	0.1840^{11}	0.0223^{11}	0.0632^{12}	0.1207^{12}	0.0331^{13}	0.1163^{11}	0.1786^{11}
17	camvi-004	0.0063^{24}	0.0697^{23}	0.2179^{22}	-	-	-	-	-	-	-
18	chosun-000	1.0000^{89}	1.0000^{80}	1.0000^{89}	-	-	-	-	-	-	-
19	chtface-002	0.0108^{41}	0.1423^{39}	0.4303^{48}	-	-	-	-	-	-	-
20	cib-000	0.0249^{63}	0.0757^{26}	0.1670^{16}	-	-	-	-	-	-	-
21	ctbcbank-000	0.0133^{50}	0.1594^{44}	0.7448^{73}	-	-	-	-	-	-	-
22	cuhkee-001	0.0041^{6}	0.0143^{5}	0.0572^{5}	0.0963^{5}	0.0143^{4}	0.0333^{3}	0.0715^{3}	0.0164^{4}	0.0652^{4}	0.1193^{4}
23	cyberlink-004	0.0061^{21}	0.0538^{18}	0.2115^{21}	-	-	-	-	-	-	-
24	dahua-004	0.0038^{4}	0.0328^{12}	0.1784^{18}	0.2026^{13}	-	-	-	0.0226^{7}	0.1186^{12}	0.1983^{12}
25	deepglint-002	0.0039^{5}	0.0077^{1}	0.0237^{1}	0.0455^{1}	0.0078^{1}	0.0141^{1}	0.0292^{1}	0.0083^{1}	0.0254^{1}	0.0513^{1}
26	deepsea-001	0.0110^{43}	0.1218^{37}	0.3094^{37}	0.3778^{17}	0.0922^{18}	0.2217^{19}	0.4469^{18}	-	-	-
27	didiglobalface-001	0.0050^{11}	-	0.0986^{9}	0.1517^{9}	0.0255^{12}	0.0515^{9}	0.0979^{8}	0.0291^{12}	0.1033^{9}	0.1558^{9}
28	dsk-000	0.1961^{77}	0.9108^{63}	0.9929^{80}	-	-	-	-	-	-	-
29	expasoft-000	0.0519^{75}	0.3186^{52}	0.6796^{70}	-	-	-	-	-	-	-
30	facesoft-000	0.0057^{16}	0.0397^{13}	0.1428^{14}	-	-	-	-	-	0.1573^{16}	-
31	fujitsulab-000	0.0180^{59}	-	0.5052^{57}	-	-	-	-	-	-	-
32	glory-002	0.0109^{42}	-	0.2729^{33}	-	-	-	-	-	-	-
33	gorilla-005	0.0117^{46}	0.1463^{41}	0.5037^{55}	-	-	-	-	-	-	-
34	idemia-005	0.0111^{44}	0.2051^{46}	0.6469^{64}	0.6968^{21}	0.1349^{19}	0.4387^{21}	-	0.2786^{21}	0.7402^{24}	0.8119^{20}
35	iit-002	0.0141^{55}	-	0.3078^{36}	-	-	-	-	-	-	-
36	imagus-001	0.0276^{65}	0.3488^{54}	0.6510^{66}	-	-	-	-	-	-	-
37	imperial-002	0.0055^{13}	0.0320^{11}	0.1350^{13}	0.1972^{12}	0.0258^{13}	0.0775^{13}	0.1556^{13}	0.0359^{14}	0.1510^{15}	0.2302^{15}
38	incode-006	0.0095^{36}	-	0.3725^{43}	-	-	-	-	-	-	-
39	innovativetechnologyltd-002	0.0251^{64}	0.2701^{49}	0.6454^{63}	-	-	-	-	-	-	-
40	innovatrics-006	0.0059^{19}	0.0543^{20}	0.2210^{23}	0.3118^{15}	0.0369^{15}	0.1109^{16}	0.1984^{15}	0.0557^{17}	0.1909^{19}	0.2764^{17}
41	intellicloudai-001	0.0095^{35}	0.1044^{36}	0.4394^{50}	-	-	-	-	-	-	-
42	intellifusion-002	0.0056^{15}	0.0539^{19}	0.1690^{17}	-	-	-	-	-	0.1822^{18}	-
43	intellivision-002	0.0463^{74}	0.5999^{58}	0.9028^{76}	-	-	-	-	-	-	-
44	intelresearch-001	0.0220^{61}	0.2254^{47}	0.6184^{61}	-	-	-	-	-	-	-
45	intsysmsu-002	0.0089^{32}	0.0827^{31}	0.3138^{38}	-	-	-	-	-	-	-

Table 3: This table summarizes False Non-Match Rate (FNMR) on unmasked and masked probe images. FNMR is the proportion of mated comparisons below a threshold set to achieve FMR=1e-05 on unmasked probe images. False Match Rate (FMR) is the proportion of impostor comparisons at or above that threshold. The red superscripts give rank over all algorithms in that column. Missing entries generally mean the algorithm was not run on that particular mask variation due to time and resource constraints. Algorithms with FTE=1.00 were not run at all.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8311

	Algorithm Name	NOT MASKED	MASKED COLOR = LIGHTBLUE, SHAPE = WIDE			MASKED COLOR = LIGHTBLUE, SHAPE = ROUND			MASKED COLOR = BLACK, SHAPE = WIDE		
	COVERAGE		LO	MED	HI	LO	MED	HI	LO	MED	HI
46	iqface-000	0.0128[48]	0.0885[33]	0.2867[34]	-	-	-	-	-	-	-
47	itmo-007	0.0098[38]	0.0840[32]	0.2685[31]	-	-	-	-	-	-	-
48	kakao-003	0.0170[58]	0.1541[43]	0.4123[46]	-	-	-	-	-	-	-
49	kedacom-000	0.0391[71]	0.3444[53]	0.6188[62]	0.6848[20]	0.2663[21]	0.5975[22]	-	-	-	-
50	kneron-005	0.0296[67]	-	0.4567[52]	-	-	-	-	-	-	-
51	lookman-004	0.0398[72]	-	0.6520[67]	-	-	-	-	-	-	-
52	luxand-000	0.2167[79]	0.9732[64]	0.9988[81]	-	-	-	-	-	-	-
53	mt-000	0.0075[28]	0.0768[27]	0.2700[32]	0.3736[16]	0.0482[16]	0.1746[17]	-	0.0749[18]	0.3084[20]	0.4239[18]
54	mvision-001	0.0137[53]	-	0.3987[45]	-	-	-	-	-	-	-
55	netbridgetech-001	0.2673[81]	0.8940[62]	0.9878[79]	-	-	-	-	-	-	-
56	neurotechnology-008	0.0100[39]	0.0794[29]	0.3450[41]	0.4460[18]	0.0818[17]	0.1834[18]	0.3127[17]	0.0953[19]	0.4893[21]	0.5472[19]
57	nodeflux-002	0.0424[73]	0.4177[56]	0.7307[72]	-	-	-	-	-	-	-
58	notiontag-000	0.6814[83]	0.9966[66]	0.9992[82]	-	-	-	-	-	-	-
59	ntechlab-008	0.0033[1]	0.0179[6]	0.0642[7]	0.1126[7]	0.0137[3]	0.0413[7]	0.0953[7]	0.0208[6]	0.0842[8]	0.1348[8]
60	paravision-004	0.0088[31]	0.0124[2]	0.0281[2]	0.0476[2]	0.0125[2]	0.0181[2]	0.0313[2]	0.0135[2]	0.0327[2]	0.0581[2]
61	pixelall-003	0.0086[30]	0.0746[24]	0.2680[29]	-	-	-	-	-	-	-
62	psl-004	0.0059[20]	0.0449[14]	0.1862[19]	-	-	0.1082[15]	0.2256[16]	0.0473[16]	0.1739[17]	0.2309[16]
63	rankone-008	0.0134[52]	0.2416[48]	0.5470[58]	0.6201[19]	0.1848[20]	0.3801[20]	0.7379[19]	0.2314[20]	0.6684[23]	0.9625[21]
64	remarkai-002	0.0073[26]	0.0685[22]	0.2352[25]	-	-	-	-	-	-	-
65	rokid-000	0.0117[45]	0.1448[40]	0.4346[49]	-	-	-	-	-	-	-
66	s1-001	0.0277[66]	0.6776[59]	0.9459[77]	-	-	-	-	-	-	-
67	scanovate-001	0.2403[80]	-	0.5973[60]	-	-	-	-	-	-	-
68	sensetime-003	0.0045[9]	0.0185[7]	0.0544[4]	0.0912[4]	0.0221[10]	0.0365[4]	0.0739[4]	0.0232[9]	0.0654[5]	0.1230[5]
69	sertis-000	0.0066[25]	0.0751[25]	0.2685[30]	-	-	-	-	-	-	-
70	shu-002	1.0000[87]	-	1.0000[87]	-	-	-	-	-	-	-
71	sjtu-002	0.0052[12]	0.0475[16]	0.1912[20]	-	-	-	-	-	-	-
72	starhybrid-001	0.0104[40]	0.1923[45]	0.5033[54]	-	-	-	-	-	-	-
73	synology-000	0.0123[47]	-	0.4459[51]	-	-	-	-	-	-	-
74	tech5-004	0.0045[8]	0.0218[8]	0.0839[8]	0.1389[8]	0.0172[6]	0.0464[8]	0.0905[6]	0.0228[8]	0.0818[7]	0.1288[7]
75	tevian-005	0.0061[22]	0.0961[35]	0.5044[56]	-	-	-	-	-	0.6178[22]	-
76	trueface-000	0.0143[56]	0.1512[42]	0.4164[47]	-	-	-	-	-	-	-
77	tuputech-000	0.2014[78]	0.8743[61]	0.9731[78]	-	-	-	-	-	-	-
78	uluface-002	0.0073[27]	0.0796[30]	0.2450[26]	-	-	-	-	-	-	-
79	upc-001	0.0167[57]	-	0.4723[53]	-	-	-	-	-	-	-
80	via-001	0.0097[37]	0.1234[38]	0.3406[40]	-	-	-	-	-	-	-
81	videmo-000	0.0140[54]	-	0.5509[59]	-	-	-	-	-	-	-
82	videonetics-002	0.6032[82]	0.9941[65]	0.9996[83]	-	-	-	-	-	-	-
83	vigilantsolutions-007	0.0194[60]	0.2849[50]	0.6839[71]	-	-	-	-	-	-	-
84	visionlabs-008	0.0034[2]	0.0139[3]	0.0579[6]	0.1014[6]	0.0154[5]	0.0412[6]	0.1004[9]	0.0187[5]	0.0664[6]	0.1284[6]
85	visteam-000	0.9960[86]	1.0000[69]	1.0000[86]	-	-	-	-	-	-	-
86	vocord-008	0.0038[3]	0.0140[4]	0.0500[3]	0.0762[3]	0.0176[7]	0.0393[5]	0.0892[5]	0.0135[3]	0.0459[3]	0.0771[3]
87	winsense-001	0.0058[17]	0.0473[15]	0.1626[15]	0.2244[14]	0.0325[14]	0.0946[14]	0.1853[14]	0.0406[15]	0.1471[14]	0.2231[14]
88	x-laboratory-001	0.0058[18]	0.0517[17]	0.2569[28]	-	-	-	-	-	-	-
89	xforwardai-000	0.0056[14]	0.0235[9]	0.1064[11]	0.1615[10]	0.0197[9]	0.0606[11]	0.1156[11]	0.0255[10]	0.1091[10]	0.1608[10]

Table 4: This table summarizes False Non-Match Rate (FNMR) on unmasked and masked probe images. FNMR is the proportion of mated comparisons below a threshold set to achieve FMR=1e-05 on unmasked probe images. False Match Rate (FMR) is the proportion of impostor comparisons at or above that threshold. The red superscripts give rank over all algorithms in that column. Missing entries generally mean the algorithm was not run on that particular mask variation due to time and resource constraints. Algorithms with FTE=1.00 were not run at all.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

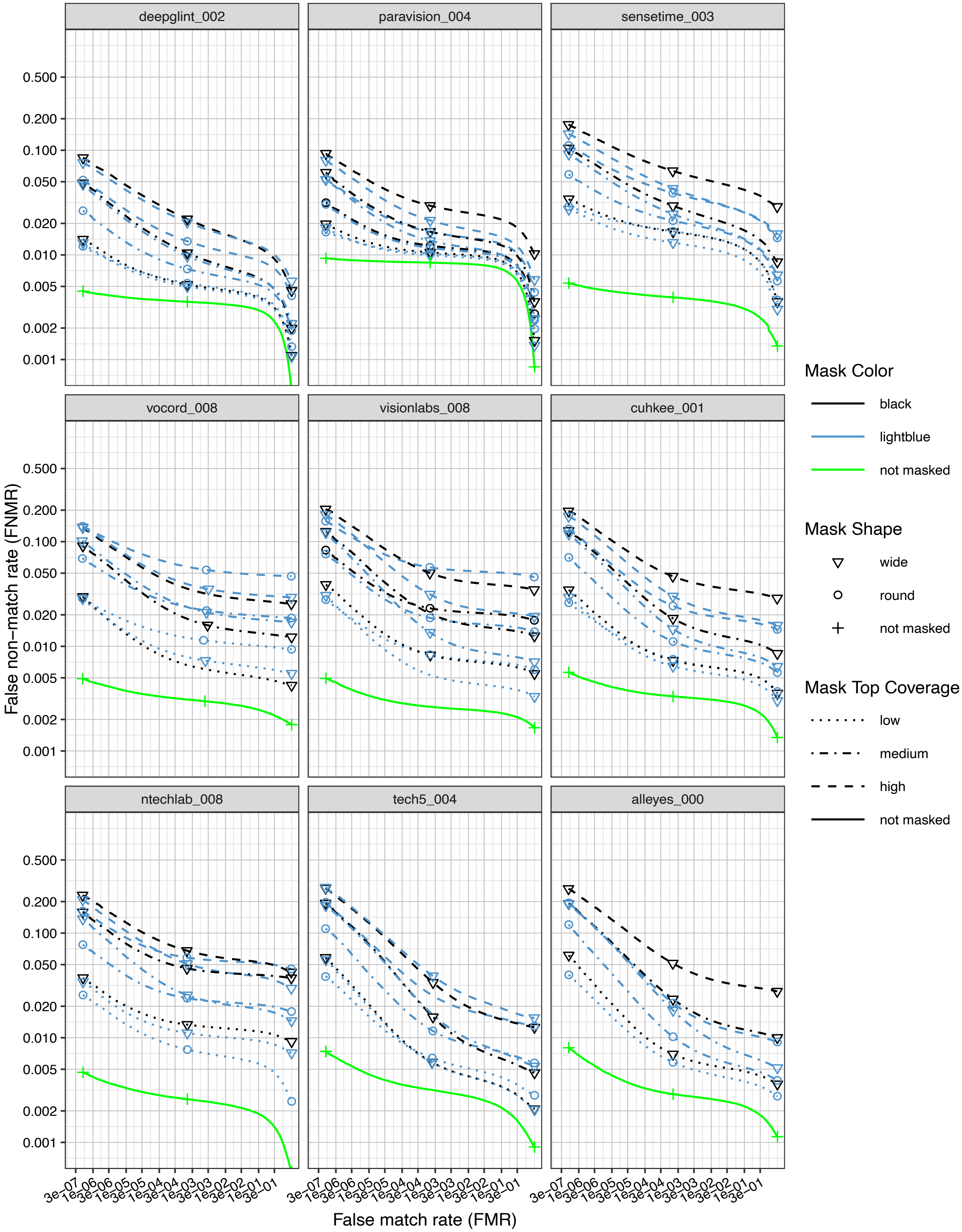


Figure 3: DET curves showing error rates on unmasked and masked images.

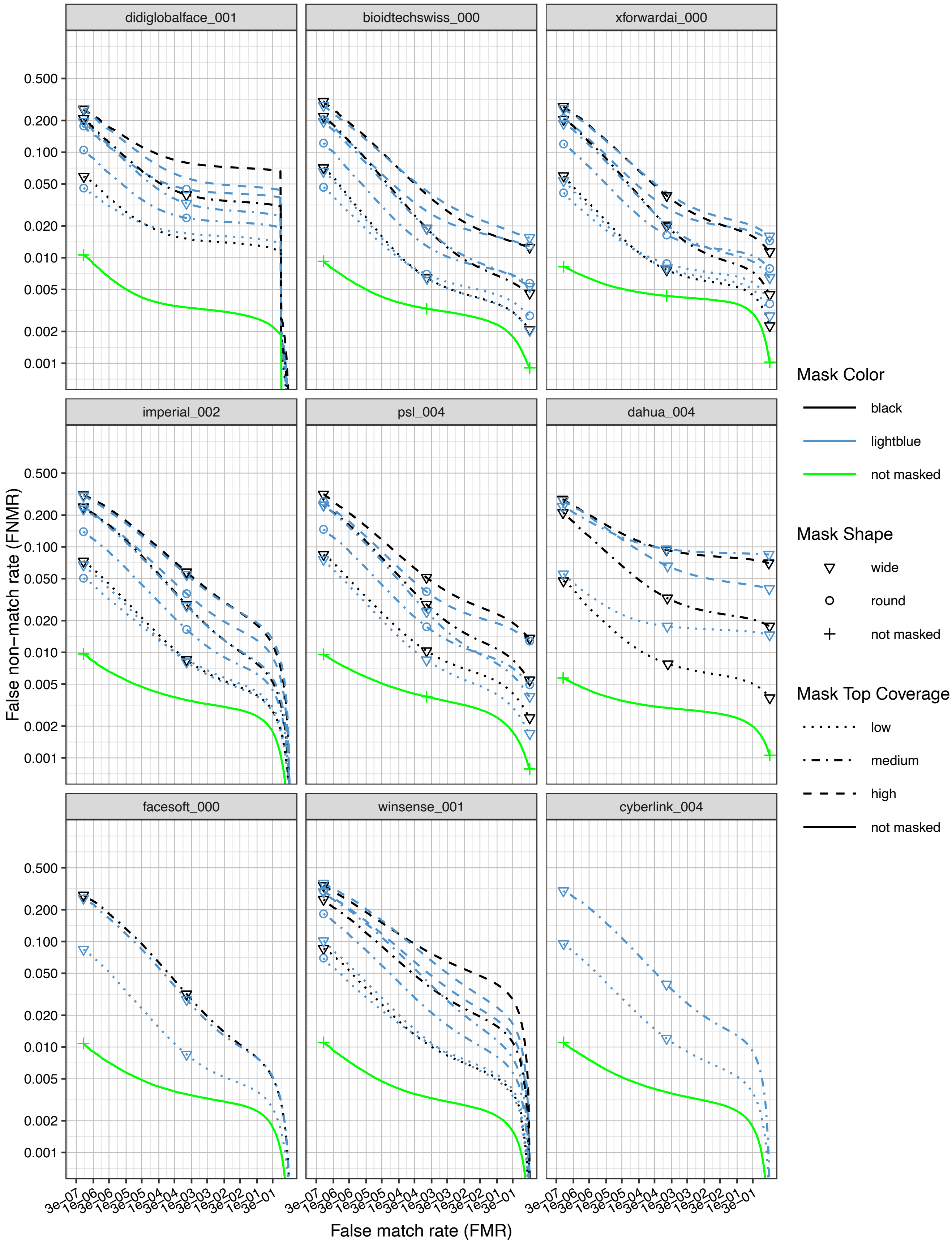


Figure 4: DET curves showing error rates on unmasked and masked images.

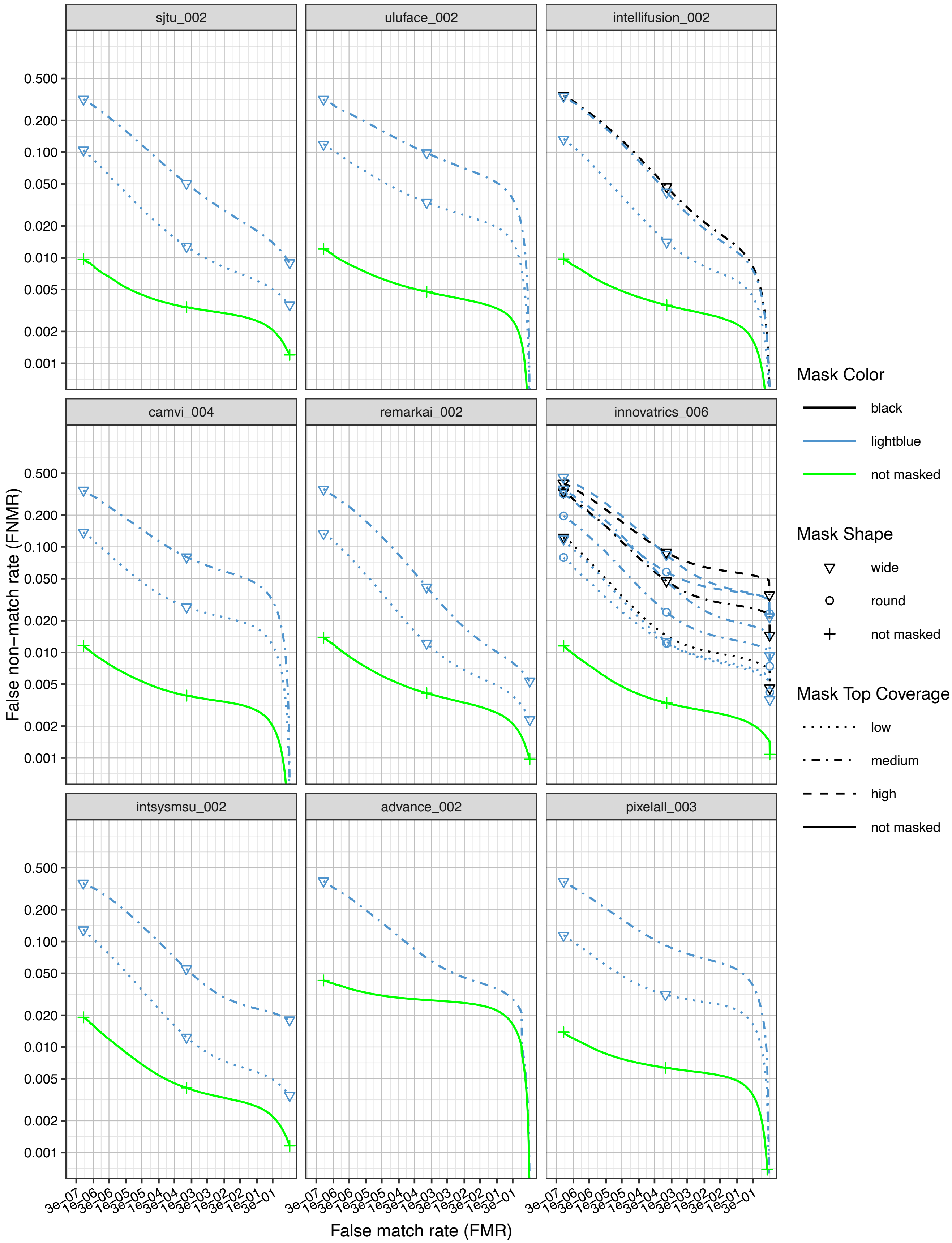


Figure 5: DET curves showing error rates on unmasked and masked images.

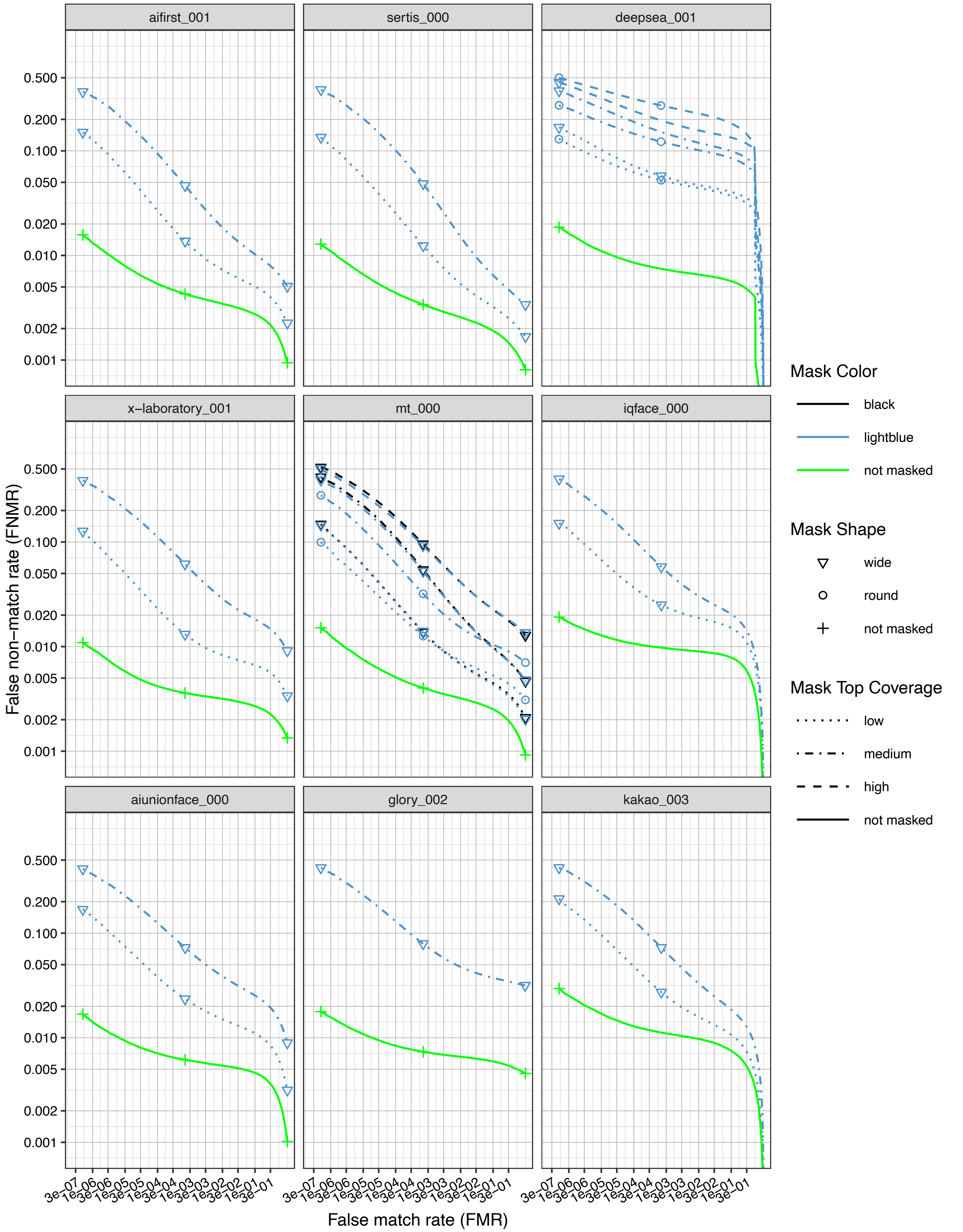


Figure 6: DET curves showing error rates on unmasked and masked images.

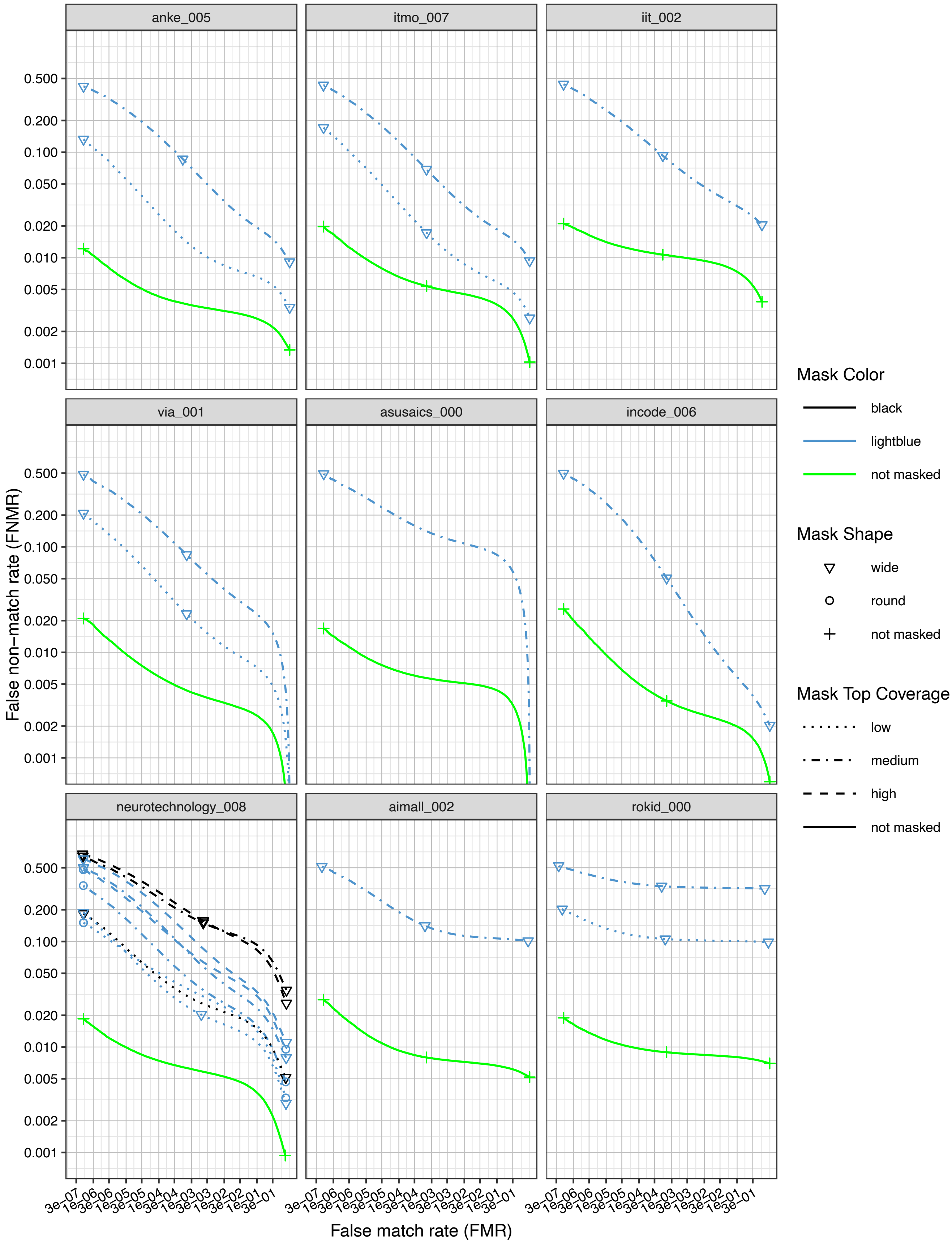


Figure 7: DET curves showing error rates on unmasked and masked images.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

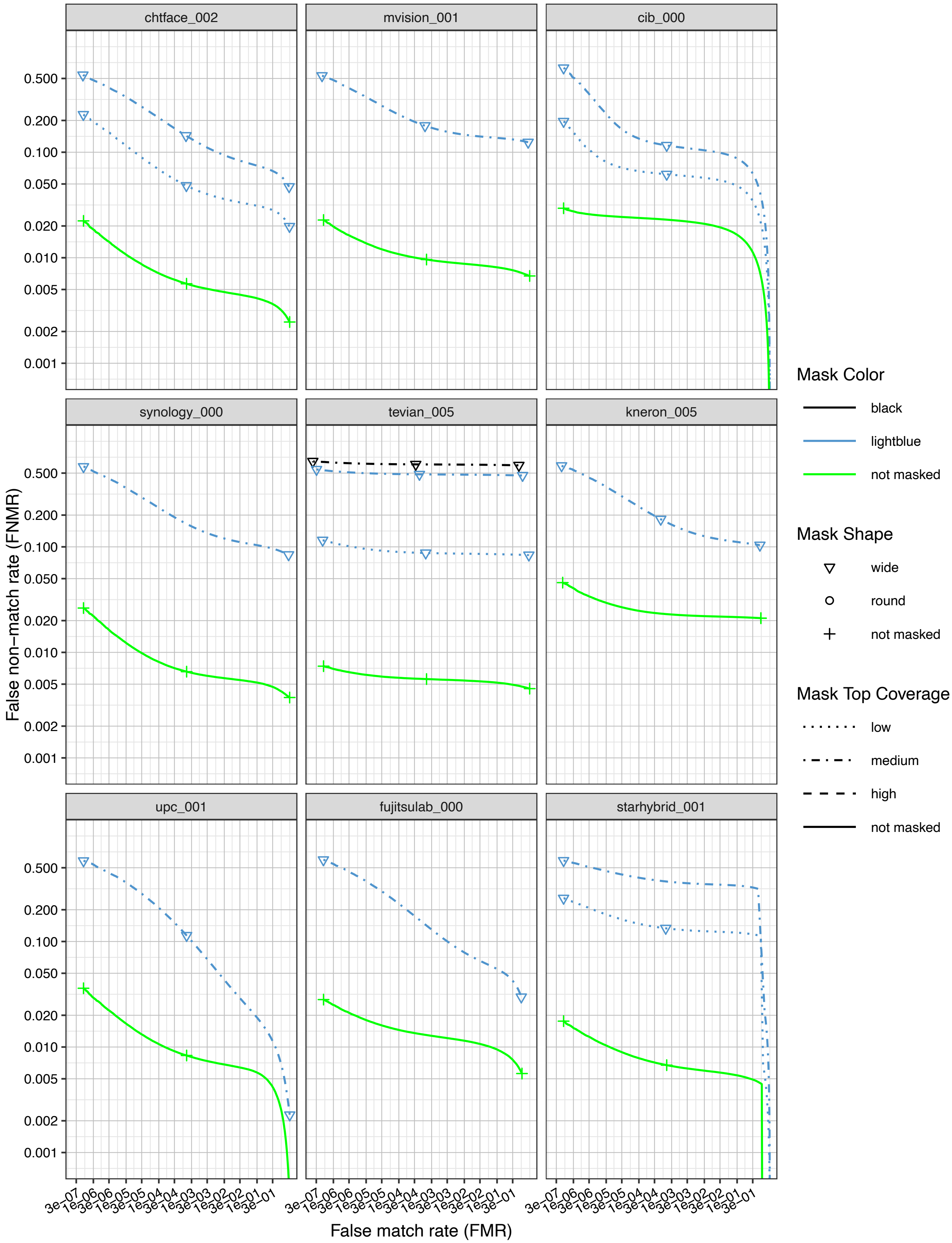


Figure 8: DET curves showing error rates on unmasked and masked images.

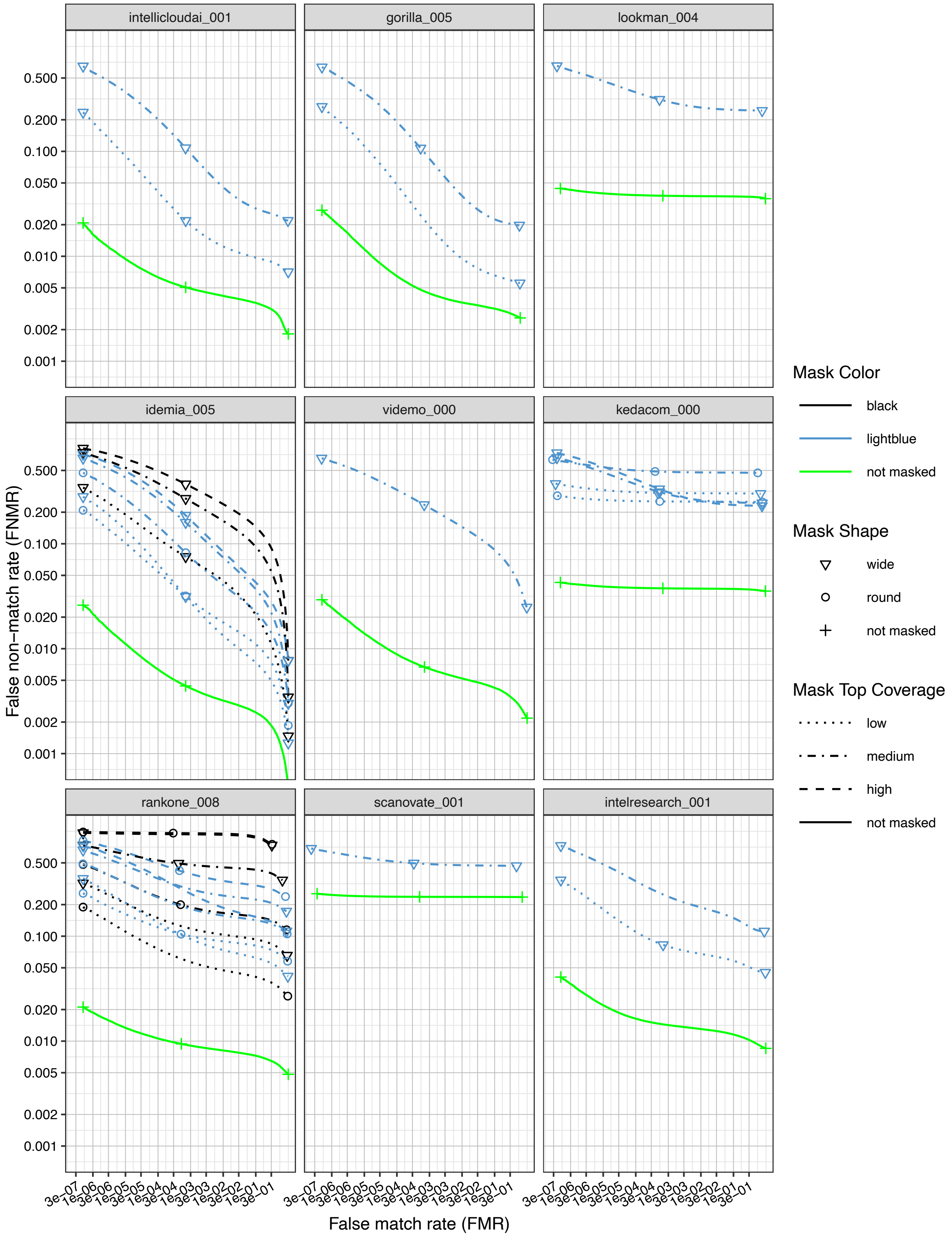


Figure 9: DET curves showing error rates on unmasked and masked images.

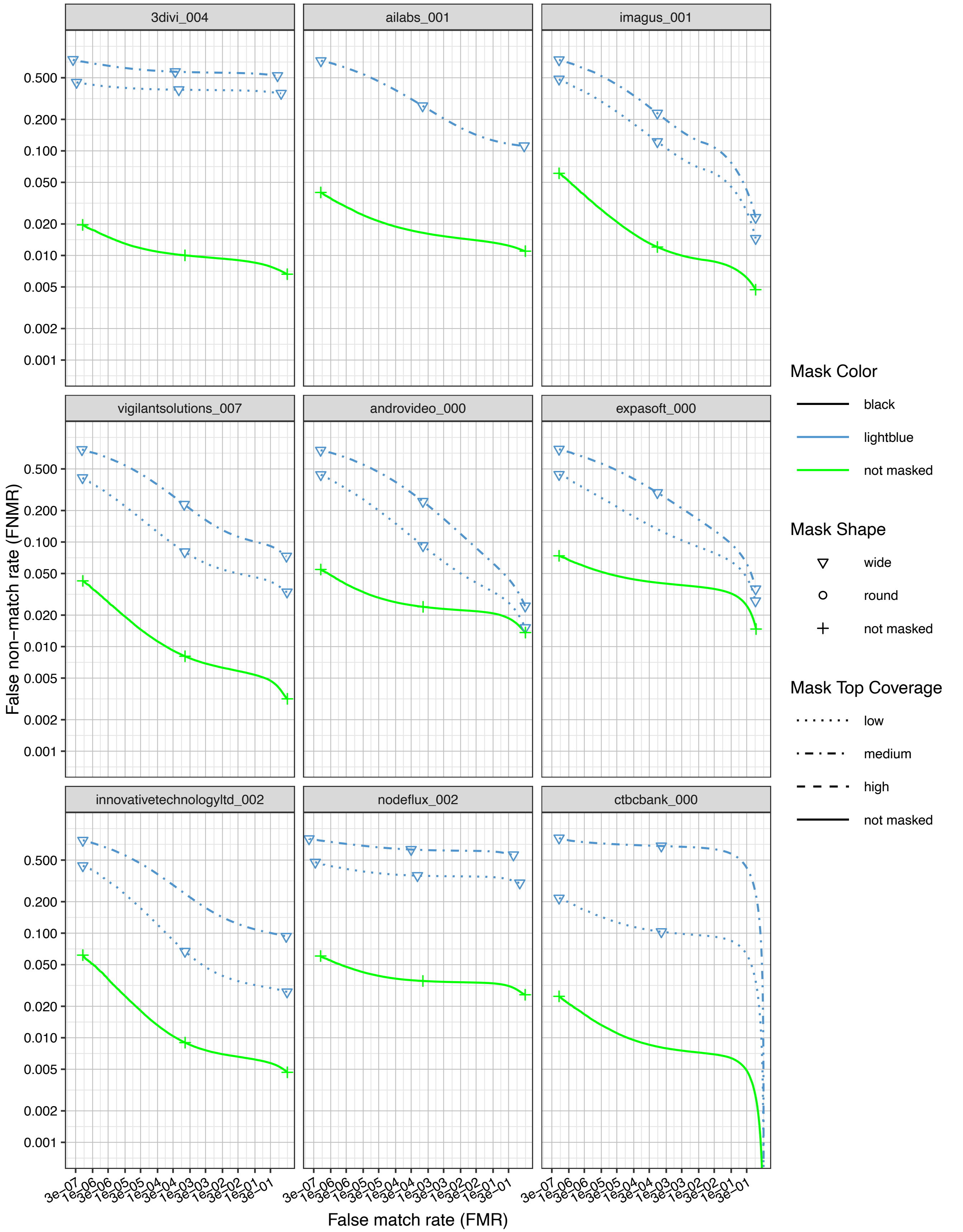


Figure 10: DET curves showing error rates on unmasked and masked images.

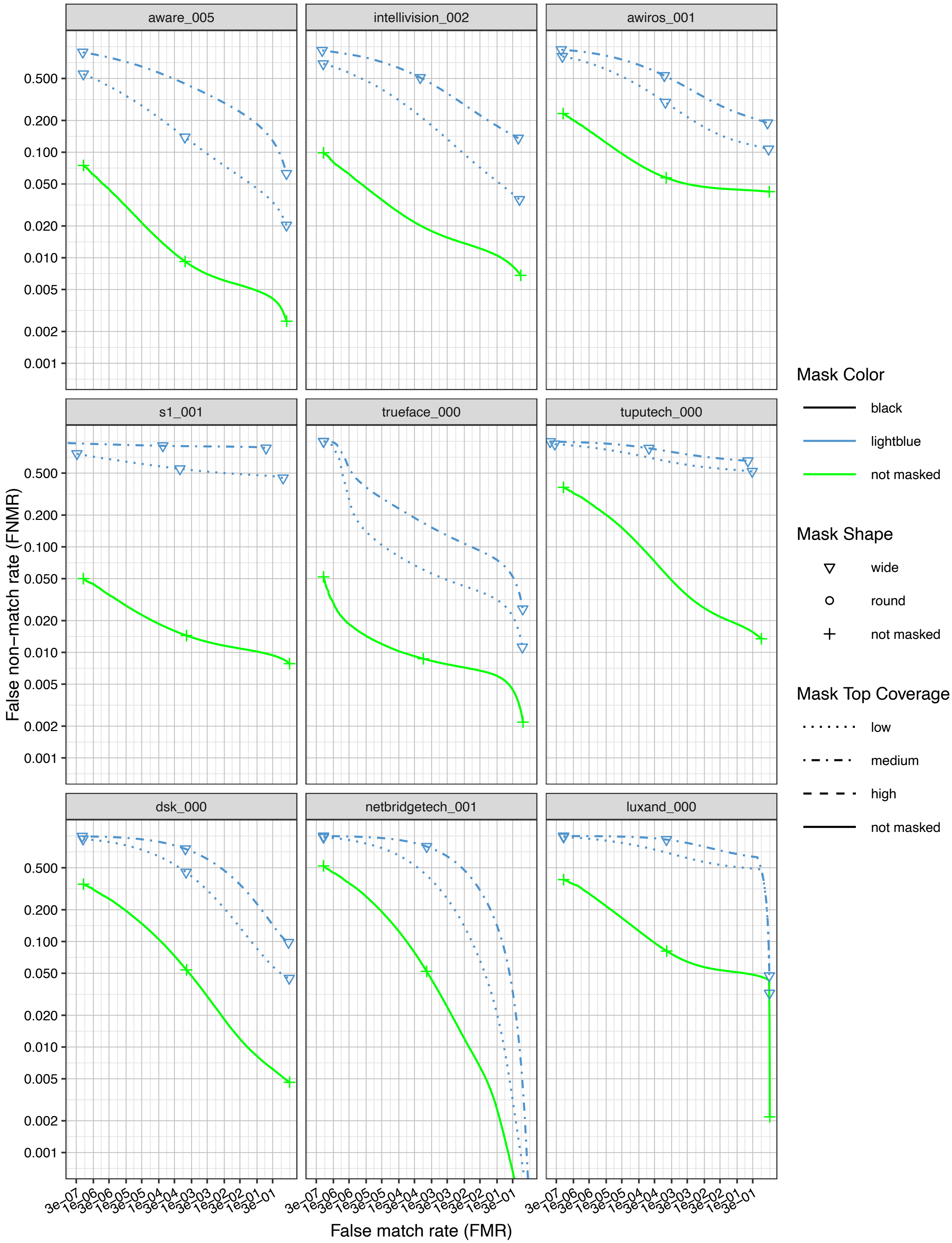


Figure 11: DET curves showing error rates on unmasked and masked images.

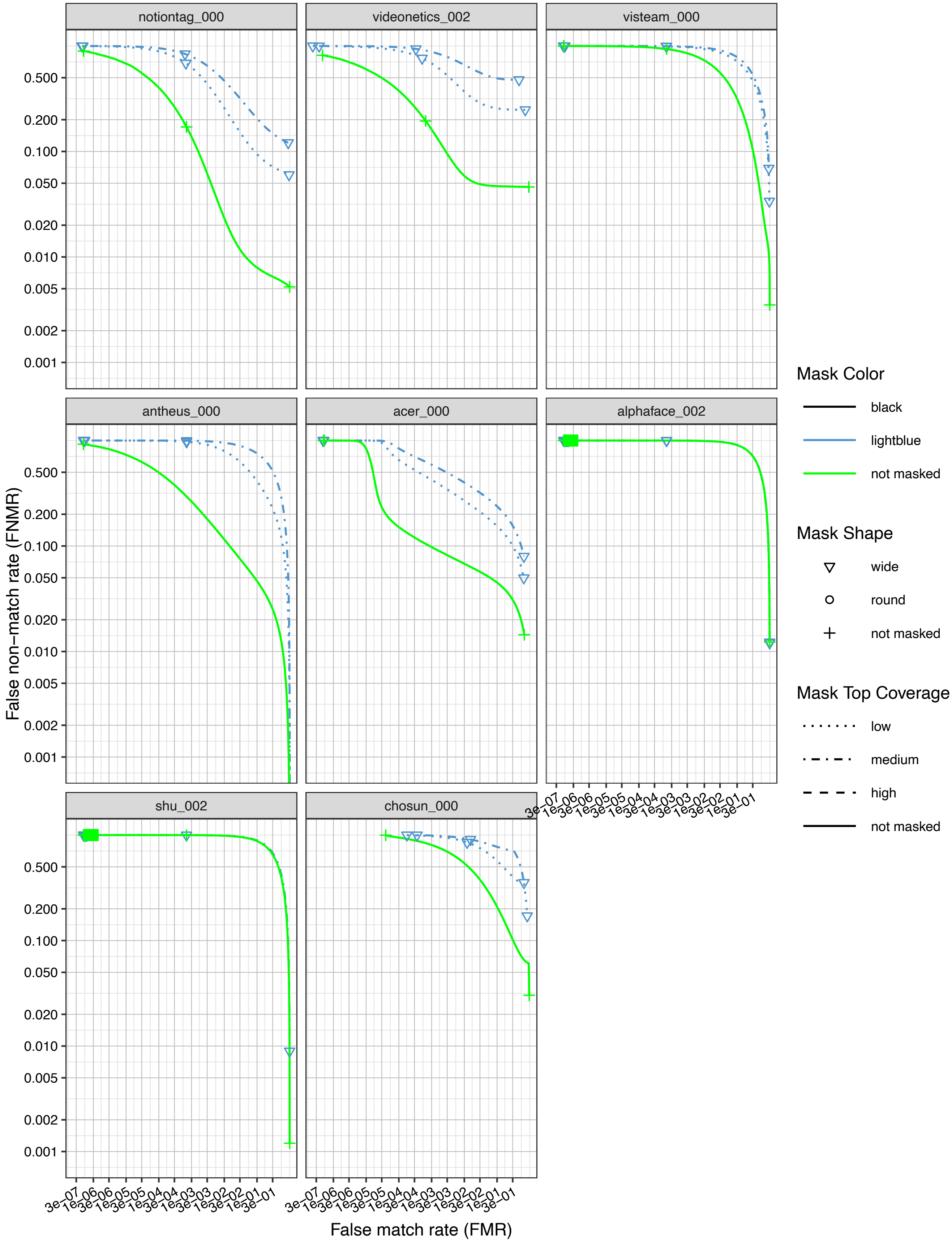


Figure 12: DET curves showing error rates on unmasked and masked images.

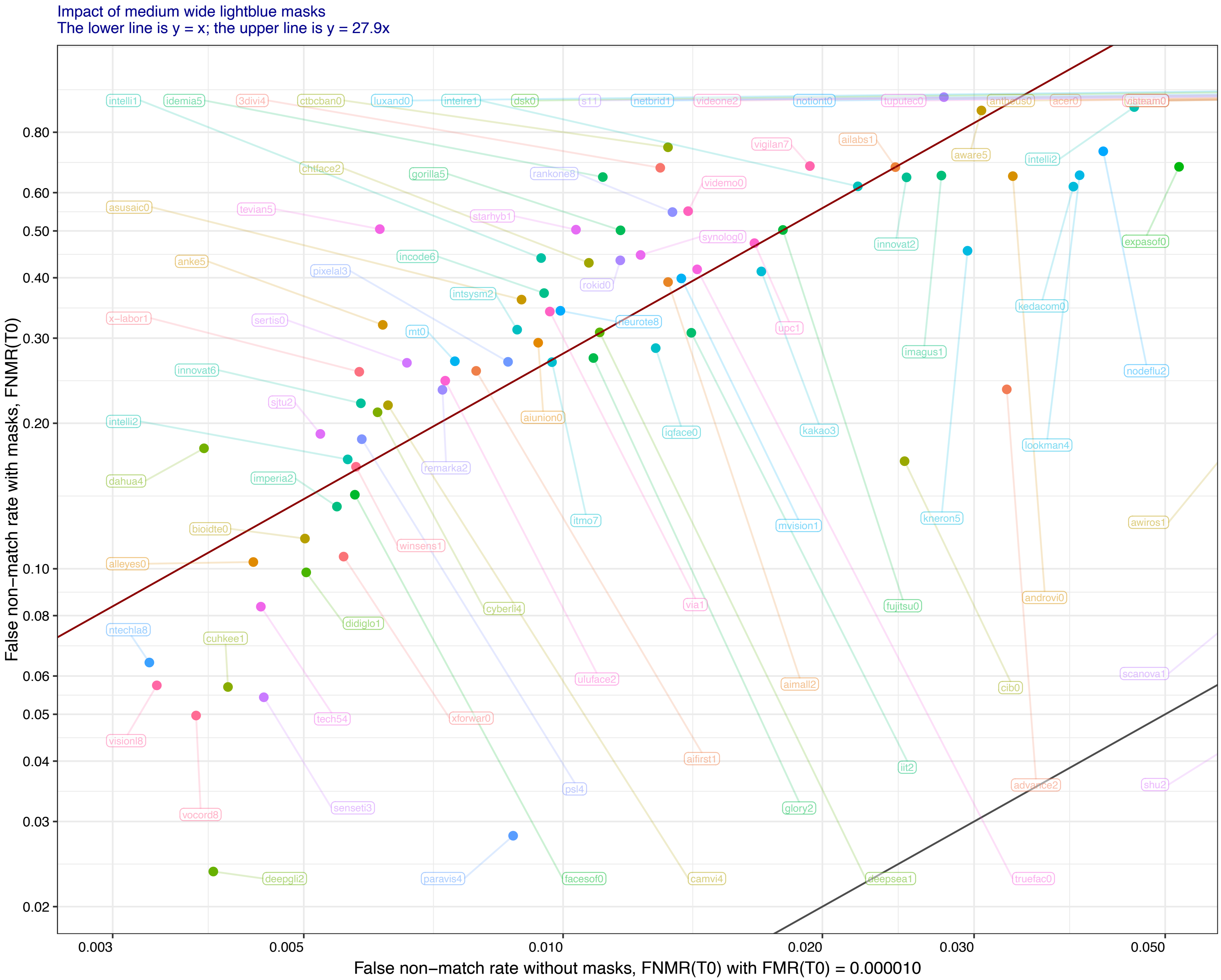


Figure 13: At a fixed threshold, a plot of FNMR with and without masks. The displacement of the red line relative to the black "parity" line shows a large increase in FNMR with masks. The value in the title is the median increase multiplier.

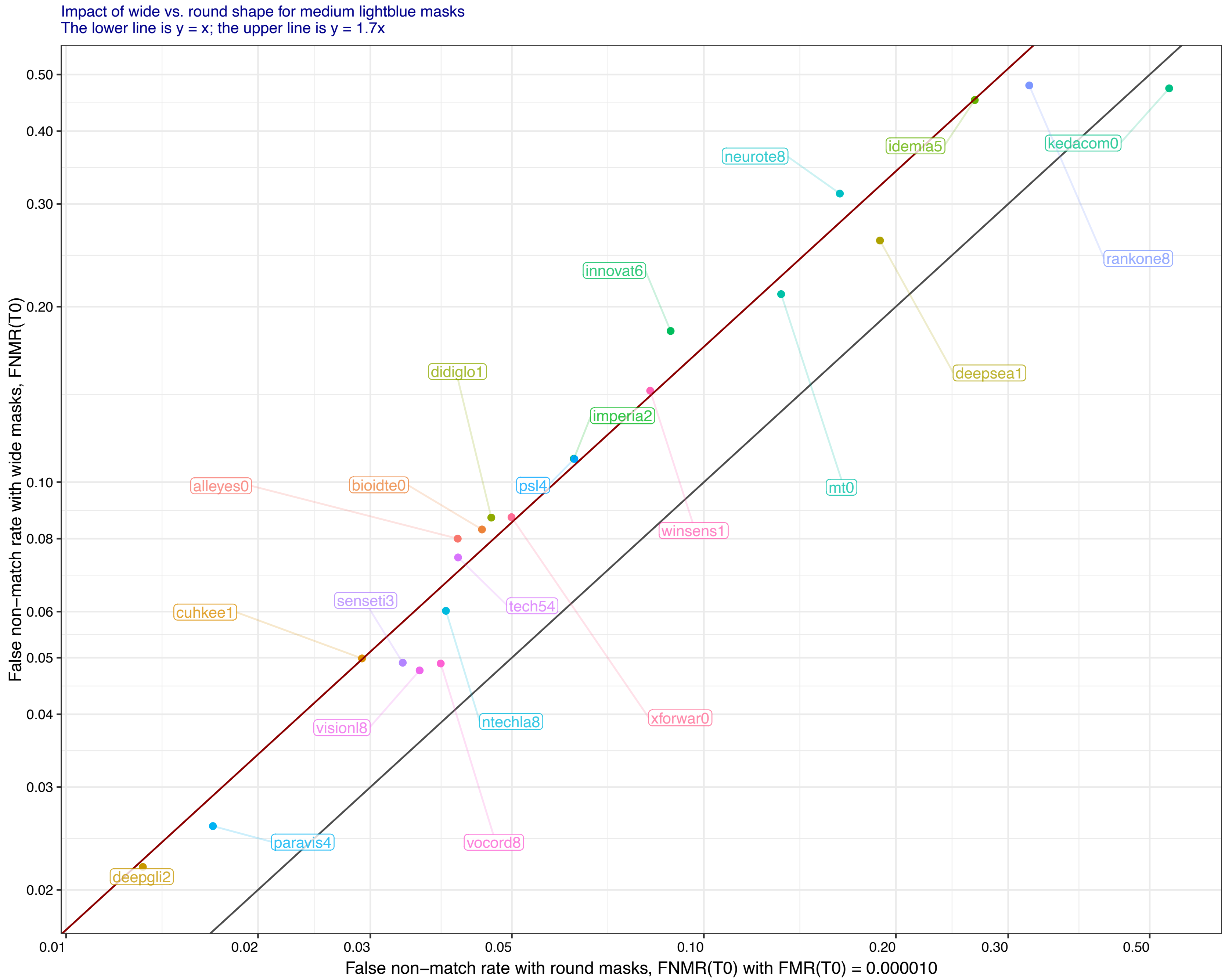


Figure 14: At a fixed threshold, a plot of FNMR with round versus wide masks. The displacement of the red line relative to the black "parity" lines shows a modest increase in FNMR with wide masks, the value in the title is the median increase multiplier.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

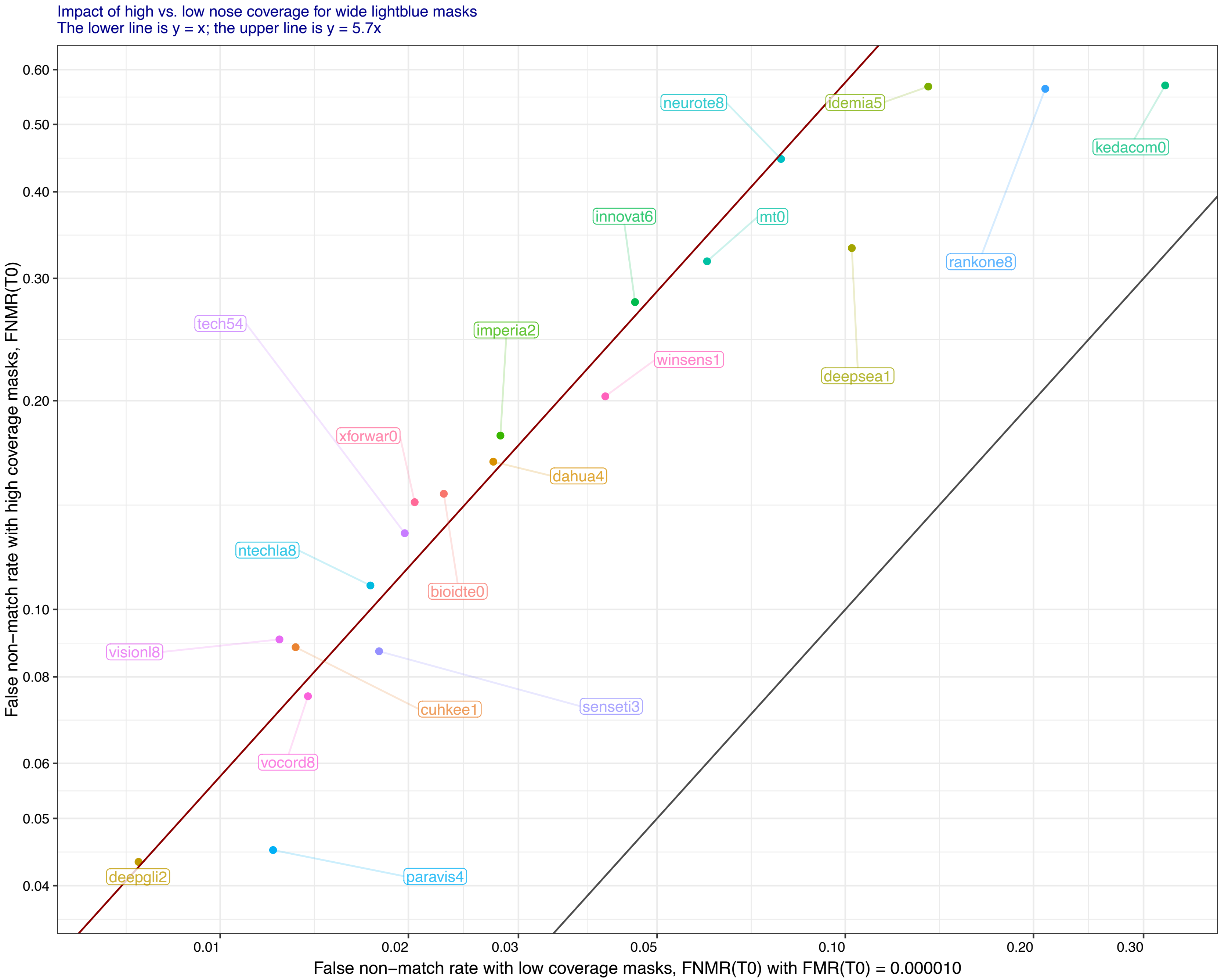


Figure 15: At a fixed threshold, a plot of FNMR with round versus wide masks. The displacement of the red line relative to the black "parity" lines shows a considerable increase in FNMR with high vs. low nose coverage masks, the value in the title is the median increase multiplier.

	Algorithm Name	COLOR = WHITE						COLOR = LIGHTBLUE						COLOR = BLACK					
		SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND		
	COVERAGE	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI
1	3divi-004	0.514	0.659	0.627	0.431	0.693	0.762	0.420	0.599	0.603	0.378	0.663	0.769	0.653	0.920	0.939	0.438	0.799	0.931
2	acer-000	0.048	0.105	0.139	0.071	0.103	0.195	0.035	0.080	0.114	0.052	0.078	0.137	0.107	0.197	0.270	0.089	0.161	0.387
3	advance-002	0.019	0.046	0.096	0.027	0.040	0.092	0.020	0.045	0.096	0.026	0.037	0.085	0.034	0.104	0.200	0.033	0.061	0.158
4	aifirst-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	ailabs-001	0.071	0.208	0.248	0.116	0.186	0.340	0.061	0.194	0.233	0.102	0.177	0.314	0.116	0.310	0.465	0.129	0.242	0.416
6	aimall-002	0.073	0.129	0.225	0.088	0.140	0.215	0.095	0.152	0.260	0.107	0.159	0.236	0.049	0.071	0.154	0.083	0.107	0.144
7	aiunionface-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	alleyes-000	0.006	0.023	0.062	0.008	0.014	0.034	0.006	0.020	0.056	0.007	0.012	0.028	0.010	0.043	0.104	0.009	0.018	0.054
9	alphaface-002	0.025	0.056	0.099	0.035	0.048	0.079	0.024	0.054	0.095	0.033	0.044	0.072	0.027	0.071	0.132	0.031	0.051	0.111
10	androvideo-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	anke-005	0.009	0.028	0.066	0.013	0.020	0.048	0.011	0.030	0.069	0.012	0.018	0.041	0.009	0.056	0.091	0.015	0.032	0.086
12	antheus-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	asusaics-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	aware-005	0.053	0.151	0.218	0.042	0.093	0.250	0.039	0.129	0.211	0.046	0.089	0.244	0.091	0.236	0.449	0.058	0.133	0.371
15	awiros-001	0.195	0.370	0.450	0.180	0.309	0.460	0.162	0.298	0.379	0.161	0.258	0.355	0.198	0.415	0.642	0.216	0.350	0.584
16	bioidtechswiss-000	0.005	0.022	0.061	0.008	0.018	0.039	0.006	0.028	0.070	0.010	0.021	0.046	0.006	0.021	0.058	0.011	0.019	0.043
17	camvi-004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	chosun-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	chtface-002	0.033	0.100	0.154	0.036	0.071	0.159	0.026	0.081	0.126	0.031	0.056	0.107	0.042	0.144	0.270	0.058	0.104	0.254
20	cib-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	ctbcbank-000	0.179	0.794	0.803	0.212	0.667	0.924	0.171	0.786	0.865	0.205	0.620	0.915	0.189	0.806	0.895	0.180	0.477	0.925
22	cuhkee-001	0.009	0.029	0.069	0.017	0.026	0.059	0.009	0.031	0.074	0.014	0.025	0.057	0.013	0.048	0.140	0.015	0.031	0.093
23	cyberlink-004	0.014	0.042	0.096	0.020	0.030	0.071	0.013	0.039	0.091	0.018	0.029	0.063	0.018	0.064	0.136	0.022	0.039	0.097
24	dahua-004	0.033	0.150	0.087	0.055	0.135	0.196	0.027	0.126	0.094	0.047	0.121	0.190	0.011	0.057	0.183	0.019	0.048	0.213
25	deepglint-002	0.002	0.009	0.028	0.003	0.005	0.014	0.002	0.012	0.031	0.004	0.006	0.017	0.003	0.010	0.024	0.003	0.006	0.018
26	deepsea-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	didiglobalface-001	0.025	0.056	0.099	0.035	0.048	0.079	0.024	0.054	0.095	0.033	0.044	0.072	0.027	0.071	0.132	0.031	0.051	0.111
28	dsk-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	expasoft-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30	f8-001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
31	facesoft-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	fujitsulab-000	0.006	0.013	0.018	0.008	0.011	0.019	0.006	0.013	0.019	0.008	0.011	0.018	0.014	0.033	0.045	0.012	0.021	0.046
33	glory-002	0.059	0.106	0.128	0.055	0.080	0.139	0.056	0.101	0.124	0.053	0.074	0.126	0.054	0.154	0.279	0.072	0.106	0.240
34	gorilla-005	0.006	0.018	0.040	0.009	0.012	0.027	0.007	0.018	0.038	0.009	0.012	0.024	0.012	0.037	0.071	0.012	0.021	0.049
35	hr-002	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
36	idemia-005	0.002	0.008	0.028	0.003	0.006	0.021	0.002	0.007	0.023	0.002	0.004	0.015	0.002	0.010	0.029	0.003	0.007	0.029
37	iit-002	0.012	0.036	0.074	0.014	0.024	0.059	0.013	0.043	0.091	0.015	0.027	0.072	0.015	0.087	0.185	0.027	0.057	0.187
38	imagus-001	0.016	0.040	0.074	0.026	0.033	0.064	0.014	0.037	0.066	0.023	0.029	0.056	0.021	0.085	0.149	0.038	0.065	0.167
39	imperial-002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	incode-006	0.002	0.008	0.020	0.002	0.003	0.008	0.002	0.008	0.018	0.002	0.003	0.007	0.002	0.012	0.031	0.002	0.004	0.012
41	innovativetechnologyltd-002	0.082	0.176	0.232	0.098	0.142	0.285	0.074	0.172	0.233	0.091	0.131	0.265	0.149	0.362	0.516	0.129	0.208	0.535
42	innovatrics-006	0.002	0.017	0.051	0.006	0.012	0.035	0.003	0.018	0.054	0.005	0.012	0.035	0.005	0.037	0.087	0.010	0.022	0.076
43	intellicloudai-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
44	intellifusion-002	0.000	0.001	0.004	0.000	0.001	0.010	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.001	0.004	0.001	0.002	0.013
45	intellivision-002	0.073	0.213	0.267	0.173	0.239	0.380	0.068	0.210	0.261	0.143	0.204	0.340	0.137	0.396	0.469	0.179	0.339	0.703

Table 5: This table summarizes Failure to Enroll (FTE) rates surveyed over 10 000 images of each mask variant. FTE is the proportion of failed template generation attempts. Failures can occur because the software throws an exception, or because the software electively refuses to process the input image as would occur if the algorithms does not detect a face or determines that the face has insufficient information. FTE is measured as the number of function calls that give EITHER a non-zero error code OR that give a "small" template containing fewer than 60 bytes. This second rule is needed because some algorithms incorrectly fail to return a non-zero error code when template generation fails but do produce a skeletal template. The effects of FTE are included in the accuracy results of this report by regarding any template comparison involving a failed template to produce a low similarity score. Thus higher FTE results in higher FNMR and lower FMR.

	Algorithm Name	COLOR = WHITE						COLOR = LIGHTBLUE						COLOR = BLACK					
		SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND		
	COVERAGE	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI
46	intelresearch-001	0.088	0.212	0.242	0.138	0.197	0.328	0.086	0.213	0.257	0.132	0.191	0.316	0.068	0.230	0.358	0.114	0.185	0.406
47	intsysmsu-002	0.008	0.055	0.117	0.021	0.041	0.120	0.007	0.047	0.110	0.015	0.033	0.100	0.036	0.105	0.231	0.040	0.075	0.218
48	iqface-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
49	isap-001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	itmo-007	0.008	0.034	0.086	0.013	0.027	0.059	0.009	0.046	0.106	0.017	0.034	0.071	0.011	0.034	0.082	0.015	0.030	0.064
51	kakao-003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
52	kedacom-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
53	kneron-005	0.063	0.184	0.206	0.106	0.163	0.307	0.058	0.166	0.212	0.094	0.146	0.276	0.101	0.440	0.505	0.154	0.325	0.574
54	lookman-004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
55	luxand-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
56	mt-000	0.005	0.021	0.061	0.011	0.022	0.047	0.006	0.024	0.063	0.011	0.021	0.045	0.007	0.023	0.059	0.011	0.021	0.046
57	mvision-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
58	netbridgetech-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
59	neurotechnology-008	0.008	0.029	0.035	0.009	0.013	0.021	0.007	0.025	0.032	0.007	0.010	0.020	0.019	0.107	0.082	0.009	0.018	0.040
60	nodeflux-002	0.402	0.598	0.538	0.449	0.635	0.835	0.440	0.671	0.628	0.482	0.681	0.877	0.602	0.835	0.915	0.418	0.604	0.927
61	notiontag-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
62	ntechlab-008	0.064	0.126	0.196	0.079	0.108	0.020	0.053	0.011	0.183	0.003	0.095	0.018	0.003	0.016	0.042	0.004	0.009	0.026
63	paravision-004	0.002	0.011	0.027	0.004	0.004	0.011	0.002	0.010	0.024	0.003	0.004	0.009	0.003	0.016	0.043	0.004	0.006	0.019
64	pixelall-003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
65	psl-004	0.004	0.017	0.042	0.009	0.018	0.038	0.004	0.015	0.037	0.007	0.014	0.029	0.011	0.028	0.058	0.018	0.034	0.078
66	rankone-008	0.136	0.414	0.293	0.180	0.276	0.459	0.117	0.358	0.292	0.154	0.229	0.386	0.153	0.470	0.770	0.109	0.230	0.770
67	remarkai-002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
68	rokid-000	0.194	0.372	0.370	0.239	0.401	0.683	0.220	0.444	0.450	0.265	0.457	0.749	0.367	0.677	0.806	0.230	0.405	0.808
69	s1-001	0.647	0.943	0.911	0.632	0.932	0.959	0.617	0.930	0.915	0.616	0.919	0.954	0.646	0.962	0.962	0.435	0.881	0.964
70	scanovate-001	0.544	0.601	0.596	0.547	0.629	0.733	0.515	0.553	0.579	0.513	0.565	0.664	0.554	0.676	0.806	0.516	0.682	0.903
71	sensetime-003	0.009	0.029	0.069	0.017	0.026	0.059	0.009	0.031	0.074	0.014	0.025	0.057	0.013	0.048	0.140	0.015	0.031	0.093
72	sertis-000	0.002	0.012	0.034	0.003	0.006	0.016	0.002	0.012	0.032	0.003	0.005	0.013	0.005	0.020	0.052	0.005	0.010	0.026
73	shu-002	0.011	0.031	0.080	0.028	0.045	0.115	0.009	0.026	0.083	0.023	0.037	0.103	0.016	0.056	0.167	0.022	0.040	0.139
74	sjtu-002	0.011	0.031	0.080	0.028	0.045	0.115	0.009	0.026	0.083	0.023	0.037	0.103	0.016	0.056	0.167	0.022	0.040	0.139
75	starhybrid-001	0.192	0.468	0.461	0.161	0.371	0.527	0.149	0.406	0.483	0.137	0.321	0.487	0.133	0.372	0.565	0.149	0.303	0.644
76	synesis-006	0.001	0.003	0.007	0.001	0.001	0.003	0.001	0.003	0.007	0.001	0.001	0.003	0.001	0.004	0.008	0.001	0.002	0.003
77	synology-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
78	tech5-004	0.005	0.022	0.061	0.008	0.018	0.039	0.006	0.028	0.070	0.010	0.021	0.046	0.006	0.021	0.058	0.011	0.019	0.043
79	tevian-005	0.125	0.463	0.370	0.181	0.271	0.581	0.148	0.650	0.557	0.208	0.359	0.705	0.131	0.786	0.787	0.122	0.272	0.758
80	trueface-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
81	tuputech-000	0.517	0.679	0.684	0.456	0.592	0.679	0.626	0.758	0.765	0.502	0.619	0.714	0.661	0.904	0.933	0.595	0.830	0.964
82	uluface-002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
83	upc-001	0.002	0.005	0.012	0.001	0.002	0.004	0.002	0.005	0.012	0.002	0.002	0.005	0.003	0.007	0.018	0.002	0.004	0.011
84	veridas-003	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
85	via-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
86	videmo-000	0.019	0.067	0.125	0.029	0.057	0.142	0.018	0.051	0.106	0.023	0.040	0.089	0.027	0.100	0.296	0.036	0.062	0.192
87	videonetics-002	0.338	0.581	0.557	0.390	0.593	0.849	0.330	0.569	0.542	0.378	0.559	0.785	0.396	0.702	0.848	0.302	0.508	0.947
88	vigilantsolutions-007	0.062	0.168	0.220	0.077	0.153	0.275	0.052	0.137	0.193	0.069	0.126	0.206	0.072	0.273	0.493	0.088	0.180	0.449
89	visionlabs-008	0.013	0.035	0.083	0.023	0.045	0.124	0.012	0.031	0.072	0.019	0.038	0.097	0.024	0.061	0.124	0.025	0.056	0.165
90	visteam-000	0.058	0.150	0.210	0.059	0.114	0.233	0.048	0.118	0.176	0.052	0.092	0.156	0.074	0.202	0.369	0.088	0.159	0.374

Table 6: This table summarizes Failure to Enroll (FTE) rates surveyed over 10 000 images of each mask variant. FTE is the proportion of failed template generation attempts. Failures can occur because the software throws an exception, or because the software electively refuses to process the input image as would occur if the algorithms does not detect a face or determines that the face has insufficient information. FTE is measured as the number of function calls that give EITHER a non-zero error code OR that give a "small" template containing fewer than 60 bytes. This second rule is needed because some algorithms incorrectly fail to return a non-zero error code when template generation fails but do produce a skeletal template. The effects of FTE are included in the accuracy results of this report by regarding any template comparison involving a failed template to produce a low similarity score. Thus higher FTE results in higher FNMR and lower FMR.

	Algorithm Name	COLOR = WHITE						COLOR = LIGHTBLUE						COLOR = BLACK					
		SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND			SHAPE = WIDE			SHAPE = ROUND		
	COVERAGE	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI	LO	MED	HI
91	vocord-008	0.013	0.046	0.087	0.025	0.047	0.096	0.011	0.052	0.089	0.031	0.059	0.111	0.009	0.050	0.093	0.018	0.037	0.095
92	winsense-001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
93	xforwardai-000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 7: This table summarizes Failure to Enroll (FTE) rates surveyed over 10 000 images of each mask variant. FTE is the proportion of failed template generation attempts. Failures can occur because the software throws an exception, or because the software electively refuses to process the input image as would occur if the algorithms does not detect a face or determines that the face has insufficient information. FTE is measured as the number of function calls that give EITHER a non-zero error code OR that give a "small" template containing fewer than 60 bytes. This second rule is needed because some algorithms incorrectly fail to return a non-zero error code when template generation fails but do produce a skeletal template. The effects of FTE are included in the accuracy results of this report by regarding any template comparison involving a failed template to produce a low similarity score. Thus higher FTE results in higher FNMR and lower FMR.

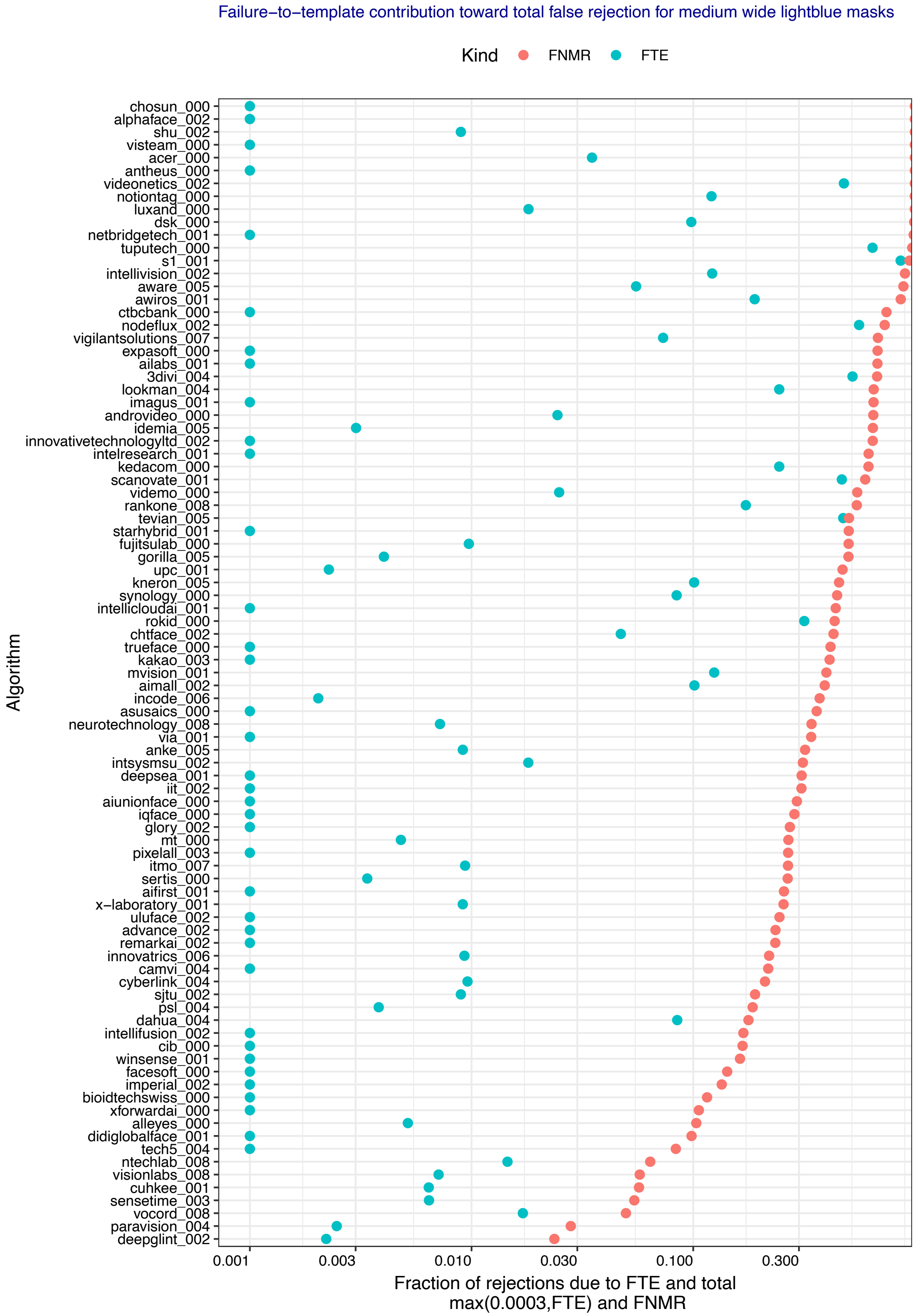


Figure 16: For each algorithm the rightmost dot shows FNMR @ FMR=0.00001 (as reported throughout this report). The left most dot shows the failure-to-template (FTE) rate over the masked verification set of 5.2M images. The gap between the two dots is attributable to low similarity score. Some FTE rates are zero - rates below 0.001 are shown as 0.001.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

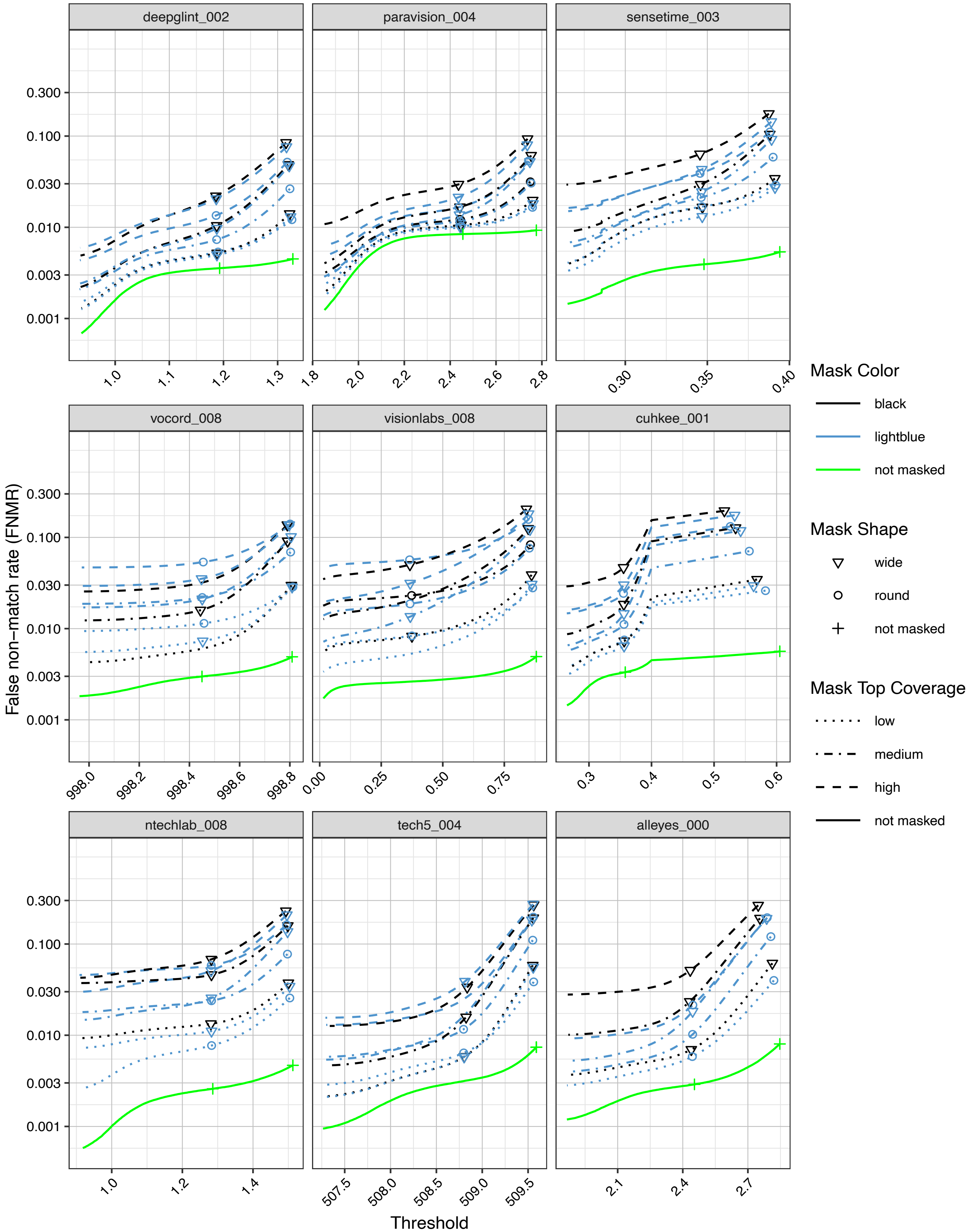


Figure 17: FNMR calibration curves on unmasked and masked images.

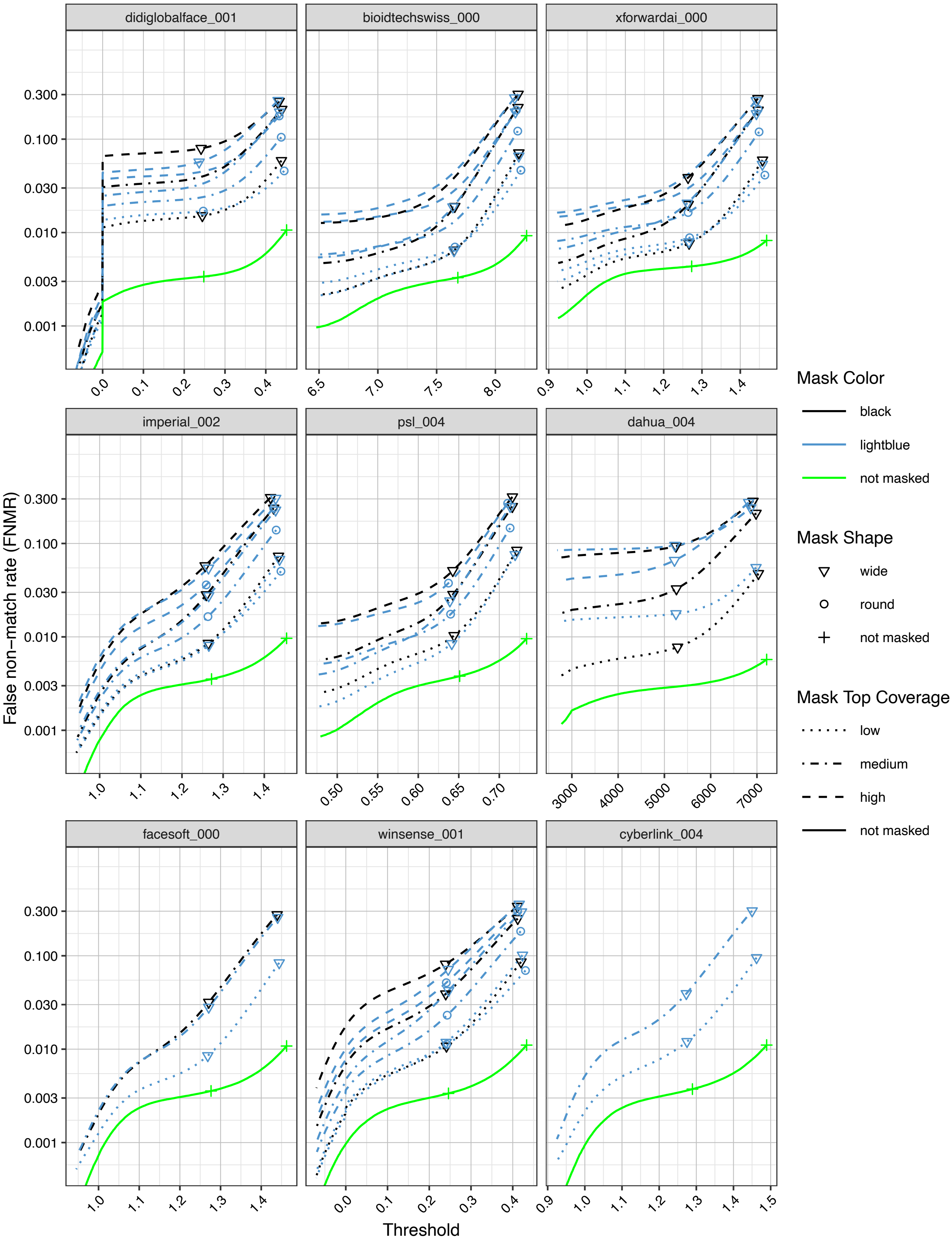


Figure 18: FNMR calibration curves on unmasked and masked images.

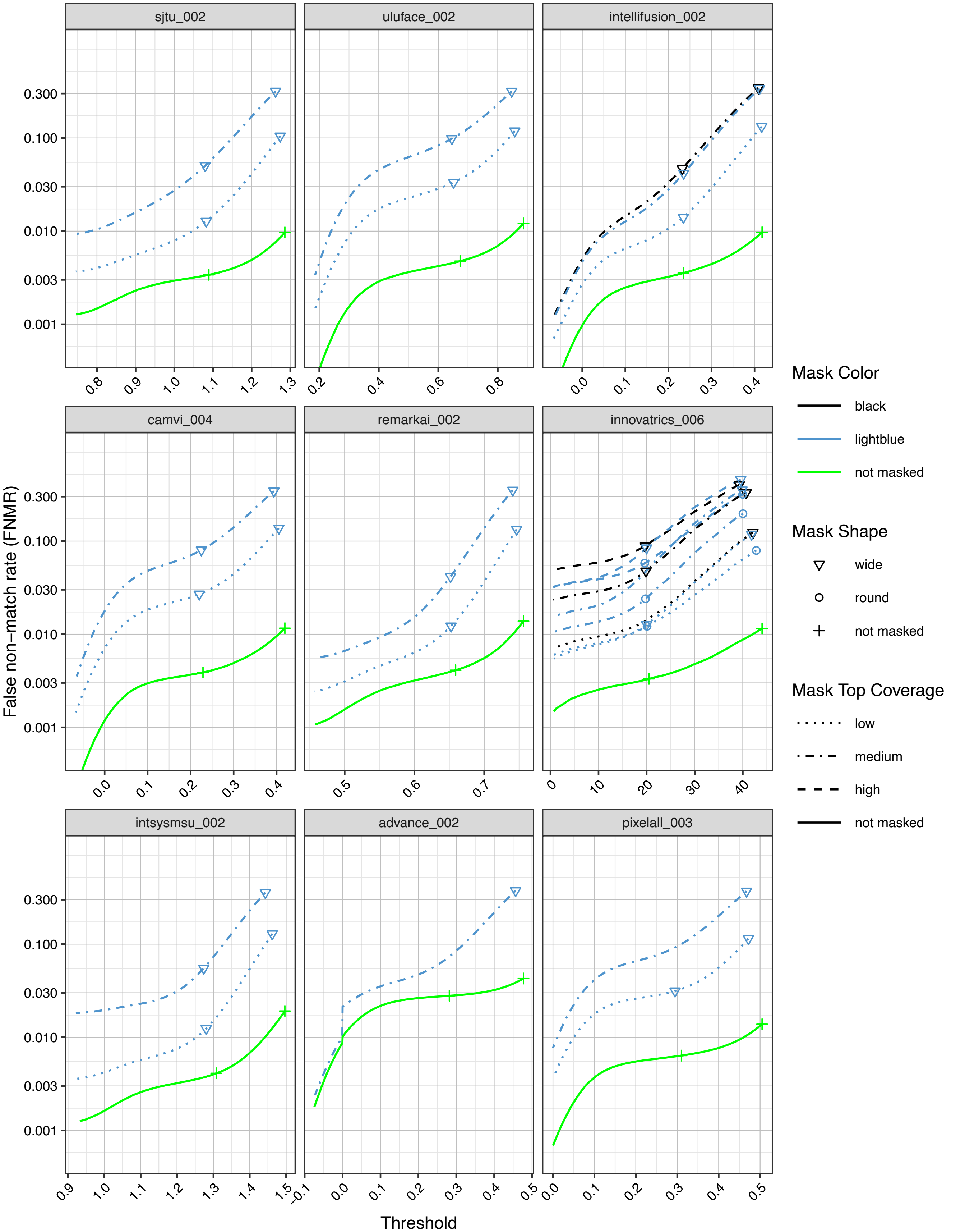


Figure 19: FNMR calibration curves on unmasked and masked images.

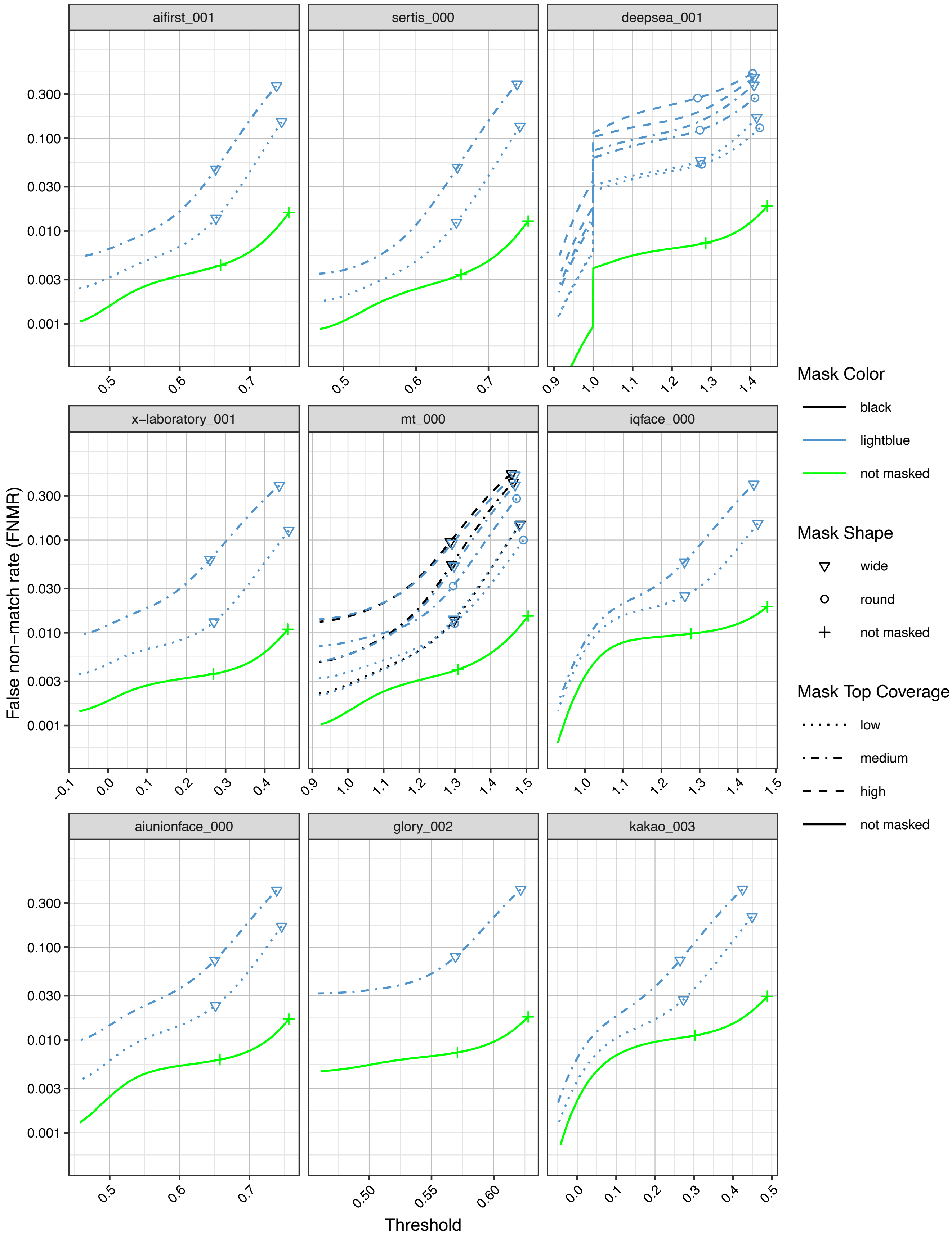


Figure 20: FNMR calibration curves on unmasked and masked images.

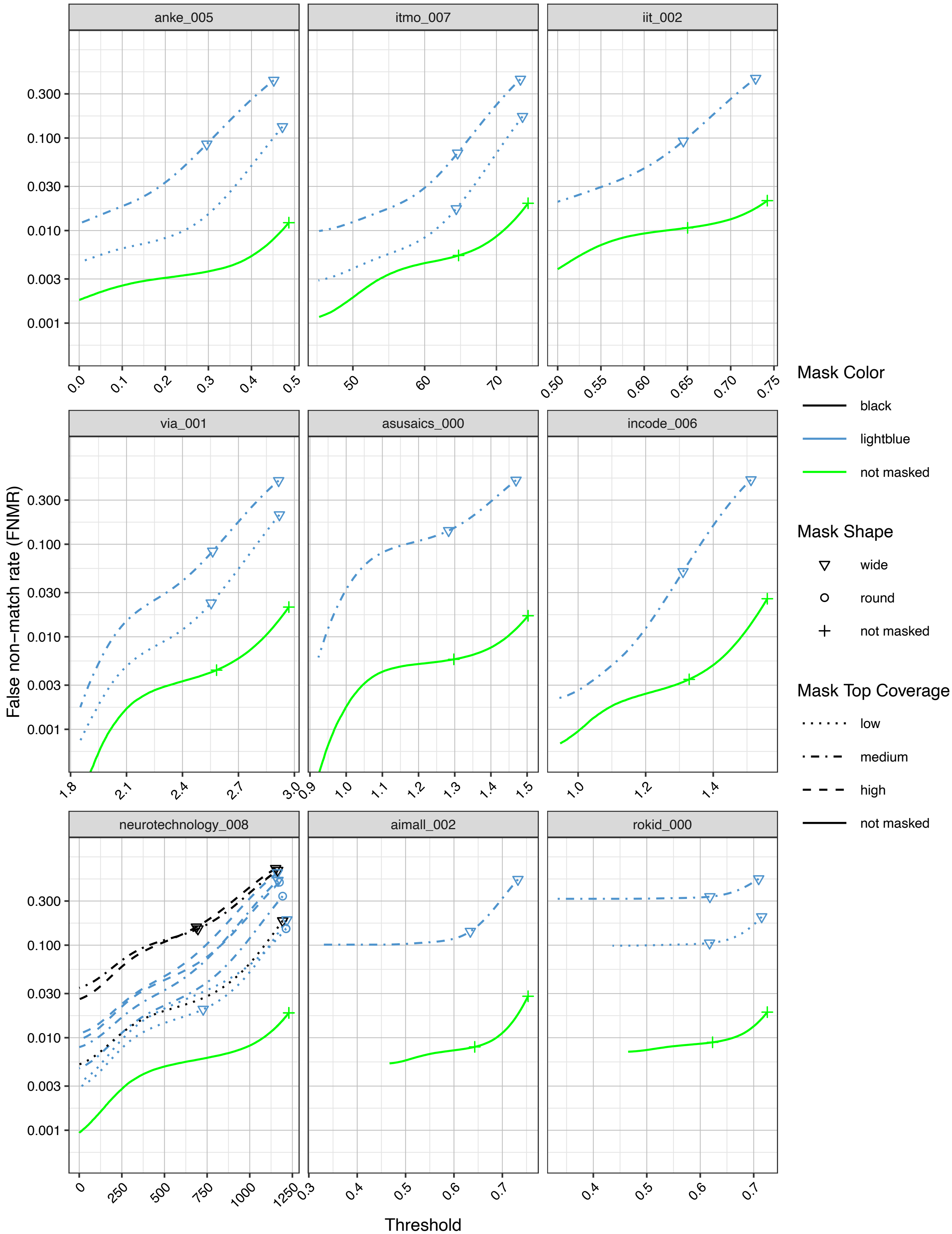


Figure 21: FNMR calibration curves on unmasked and masked images.

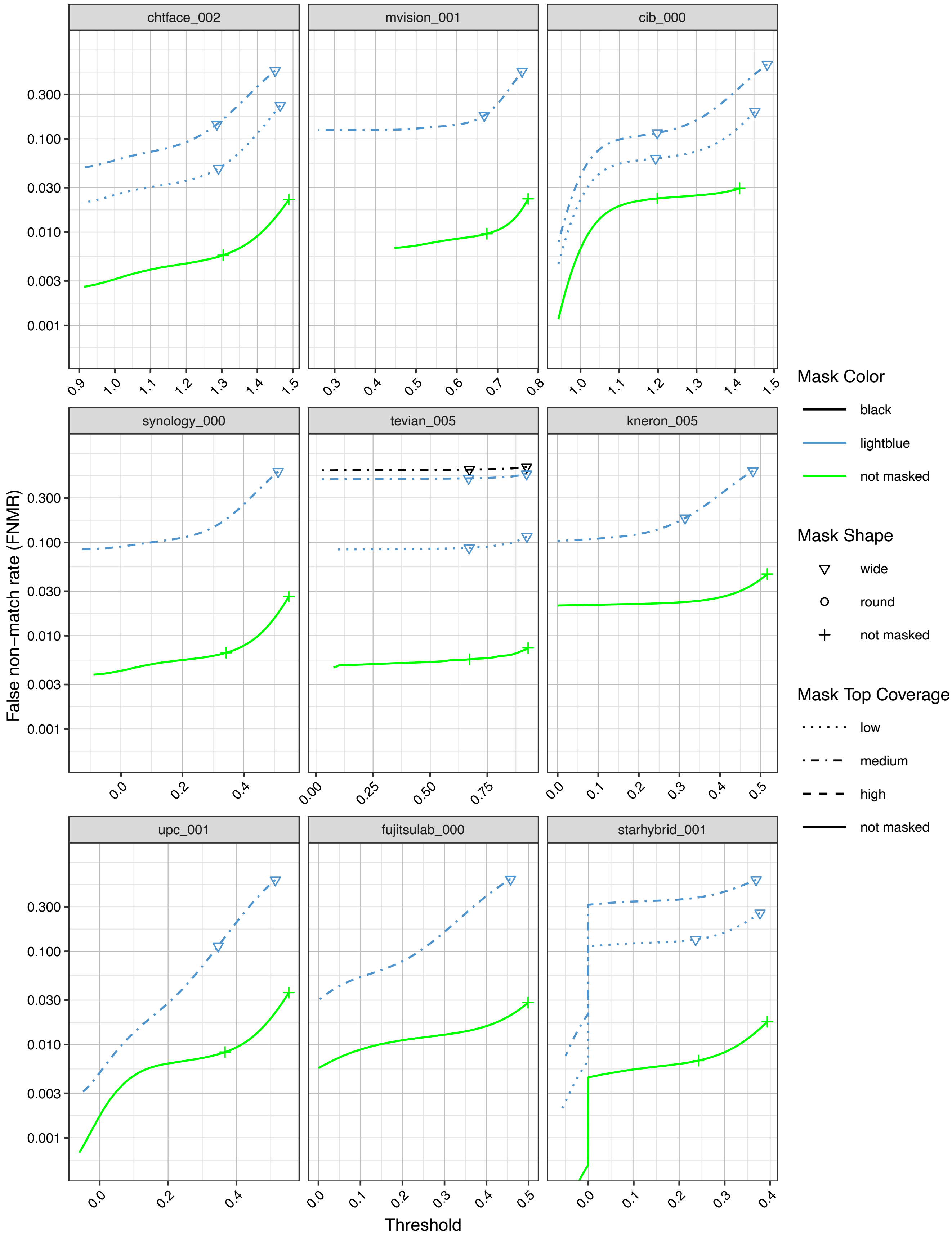


Figure 22: FNMR calibration curves on unmasked and masked images.

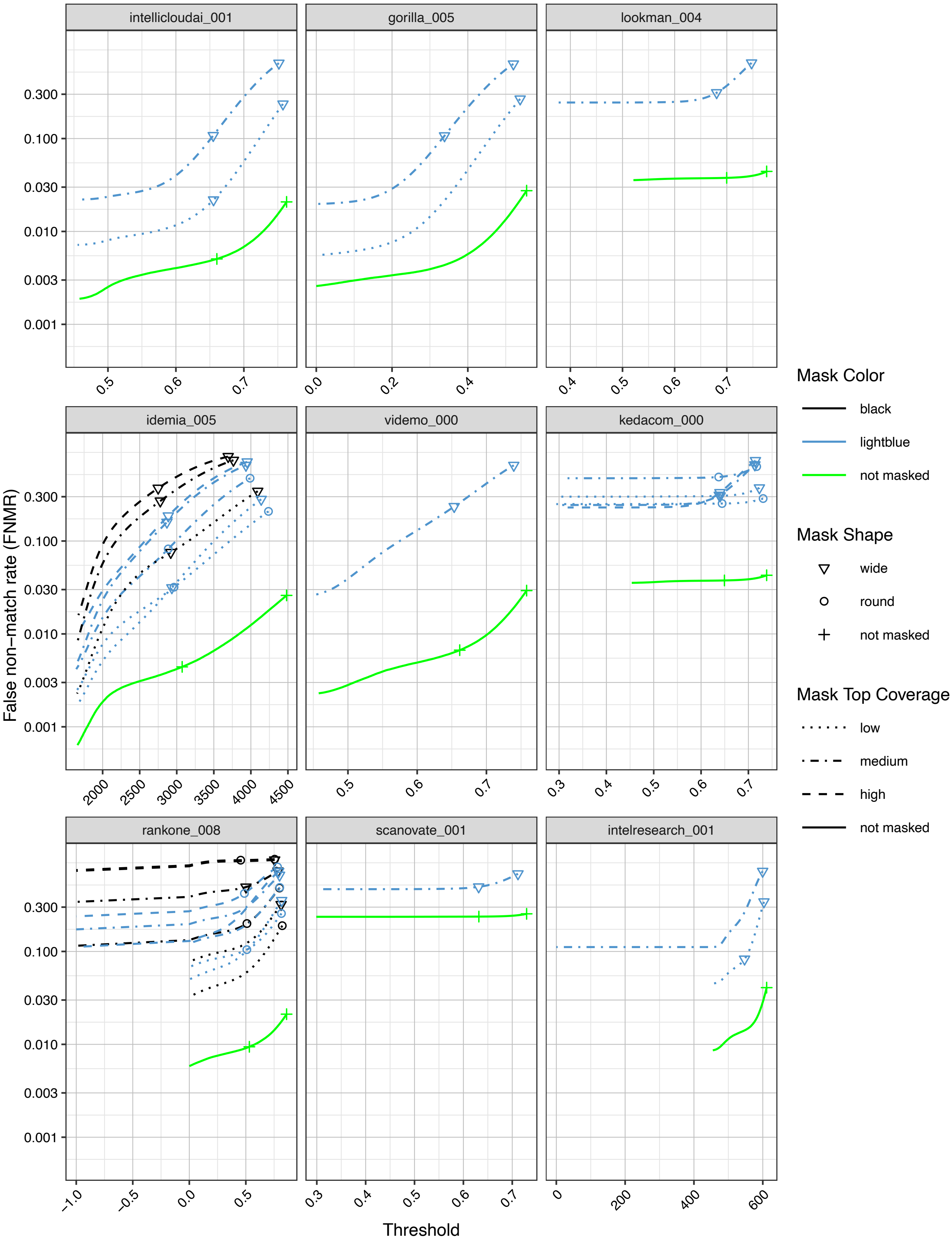


Figure 23: FNMR calibration curves on unmasked and masked images.

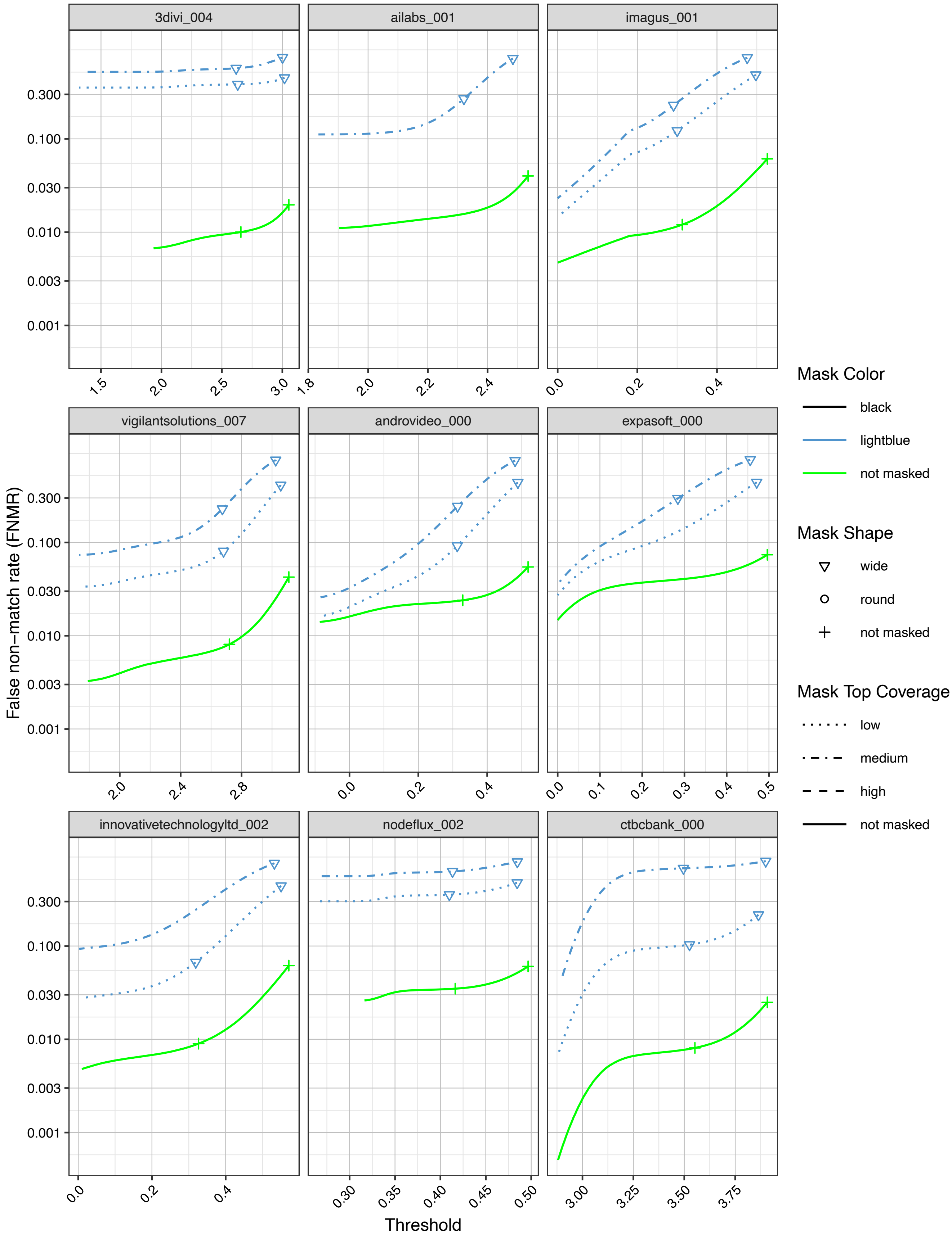


Figure 24: FNMR calibration curves on unmasked and masked images.

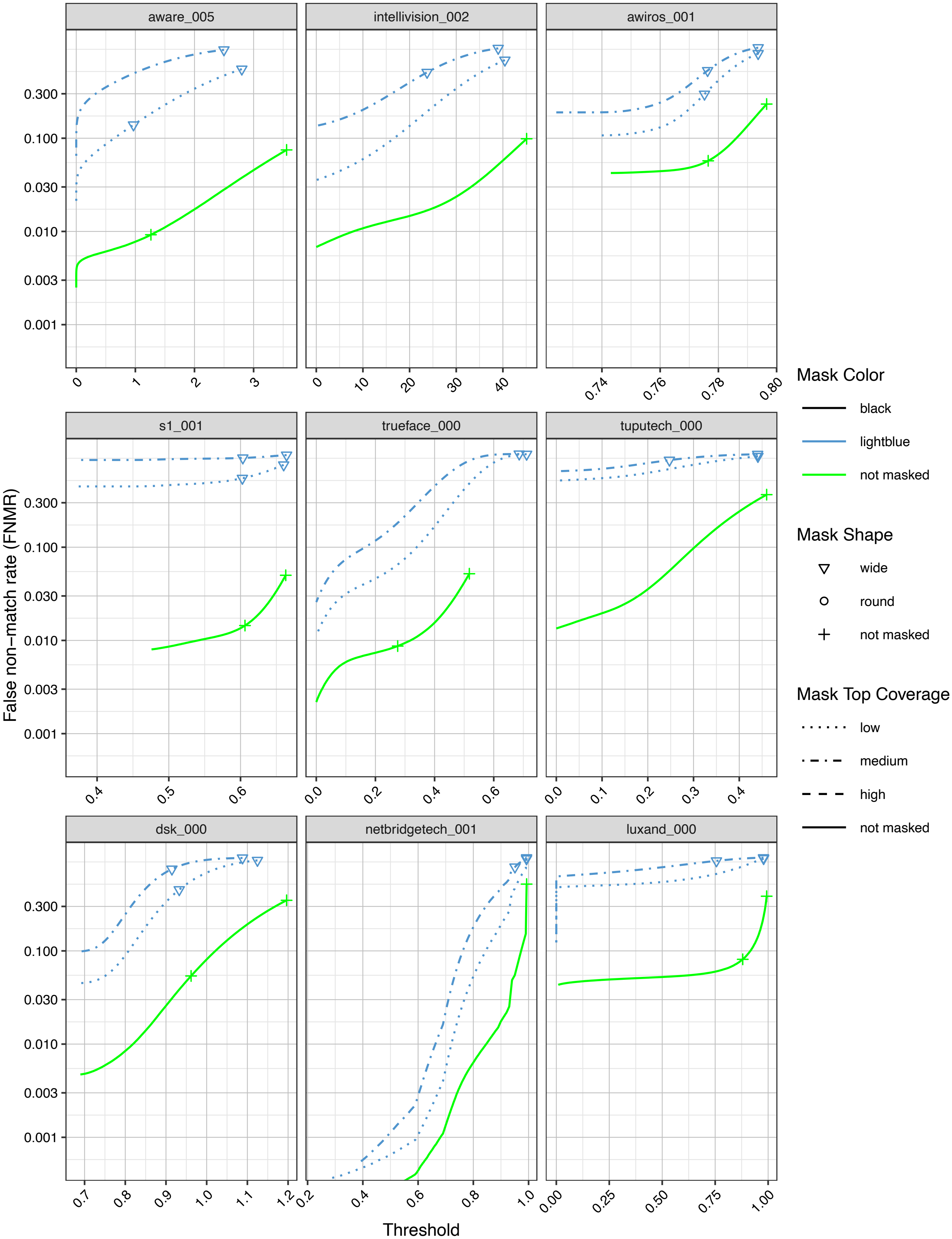


Figure 25: FNMR calibration curves on unmasked and masked images.

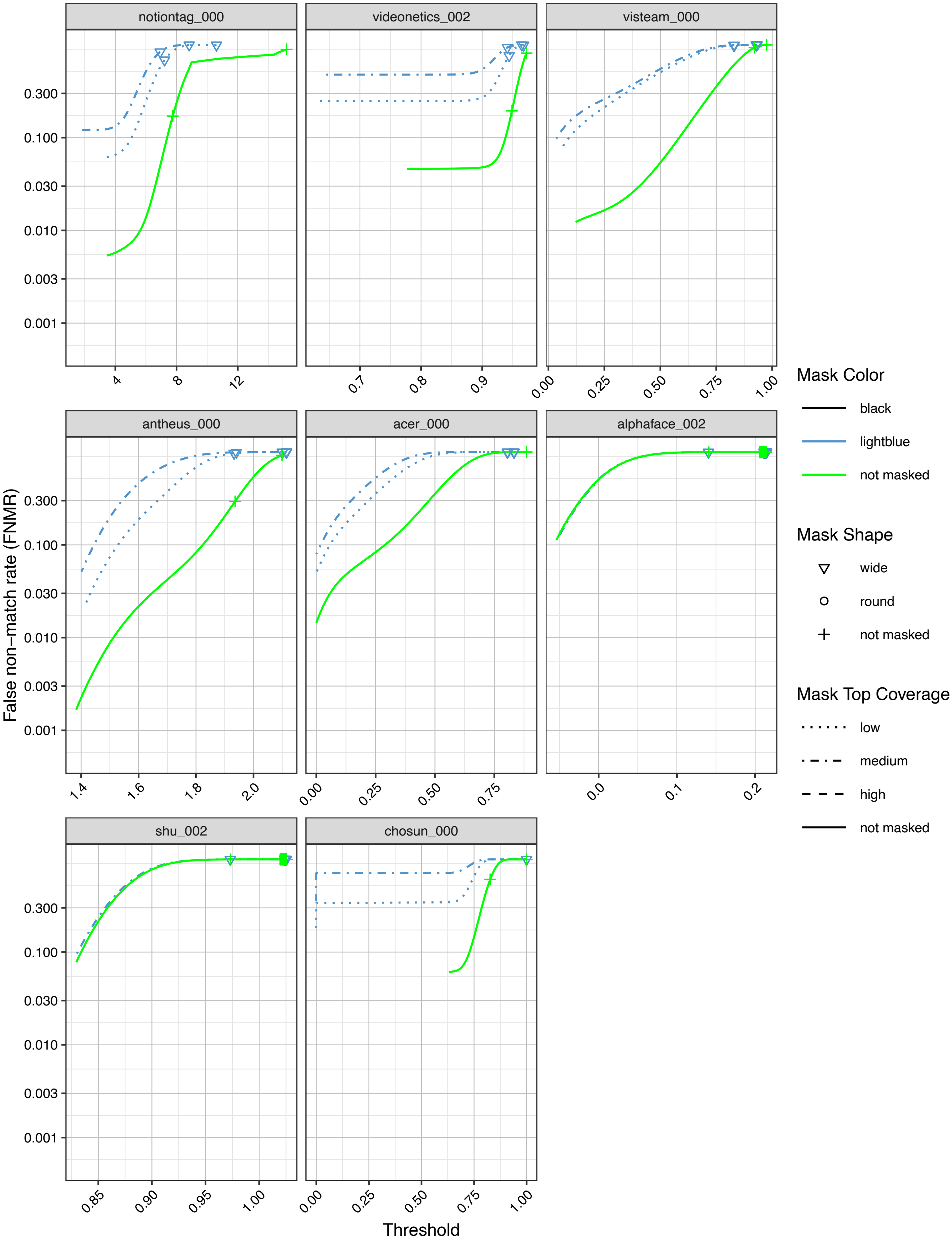


Figure 26: FNMR calibration curves on unmasked and masked images.

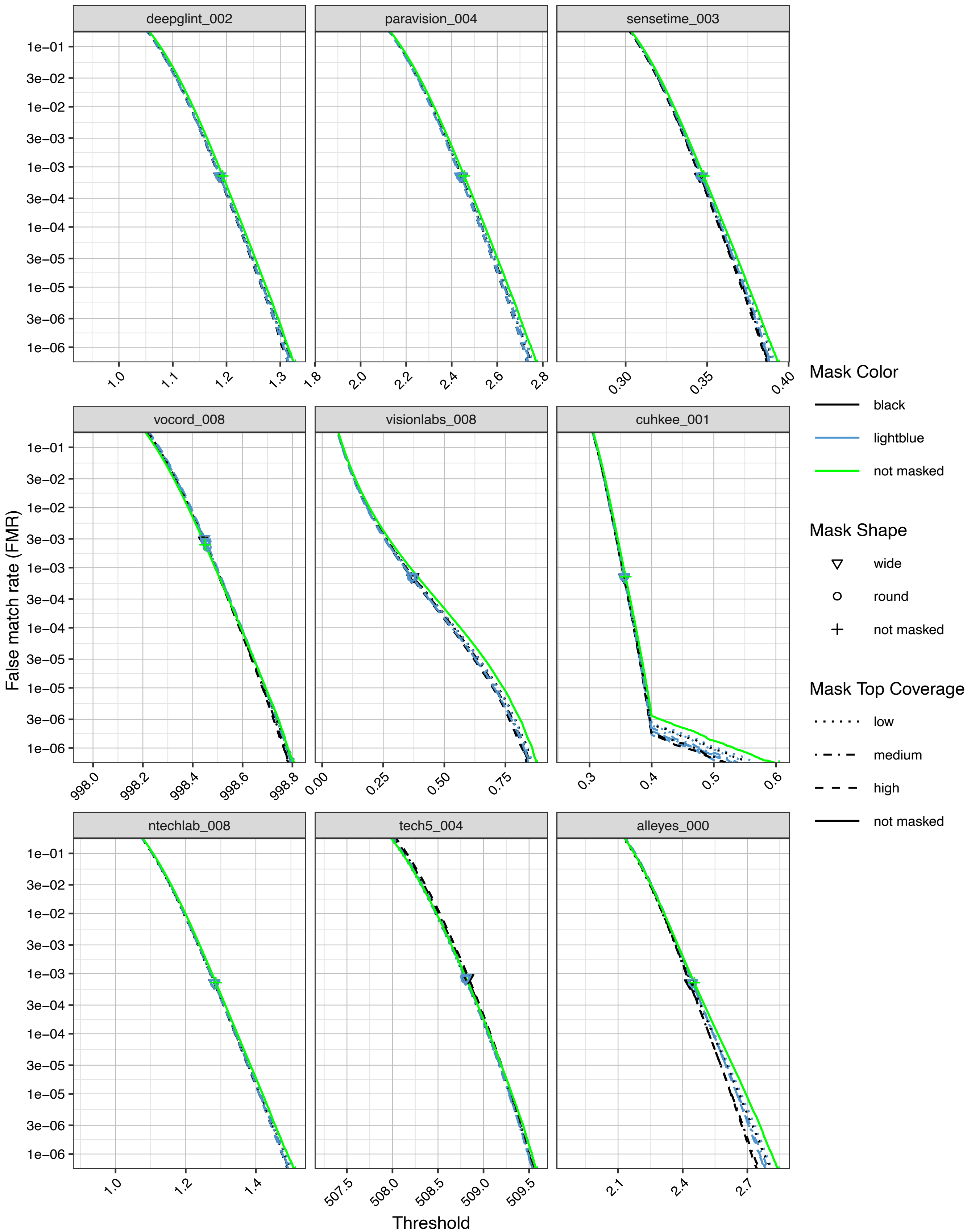


Figure 27: FMR calibration curves on unmasked and masked images.

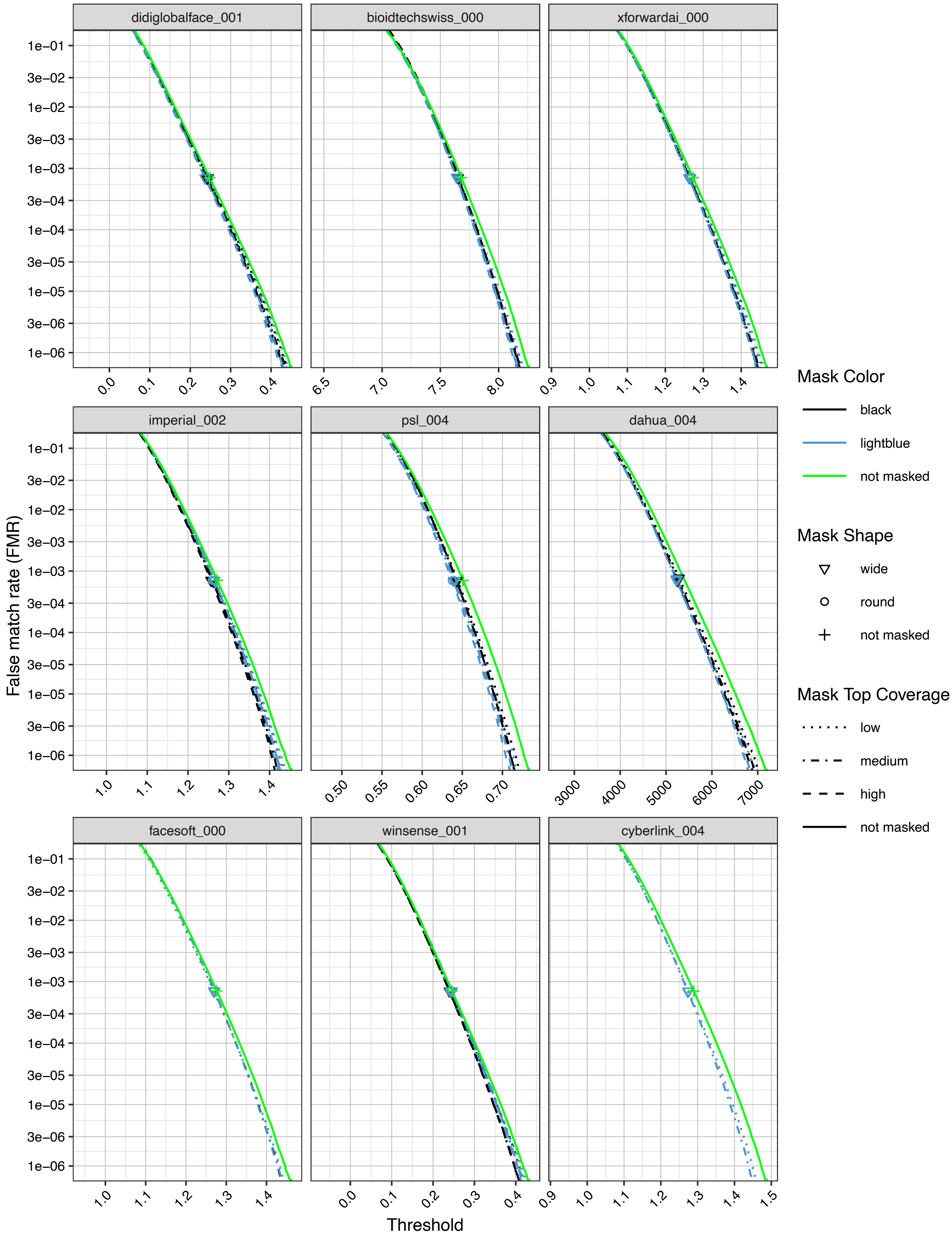


Figure 28: FMR calibration curves on unmasked and masked images.

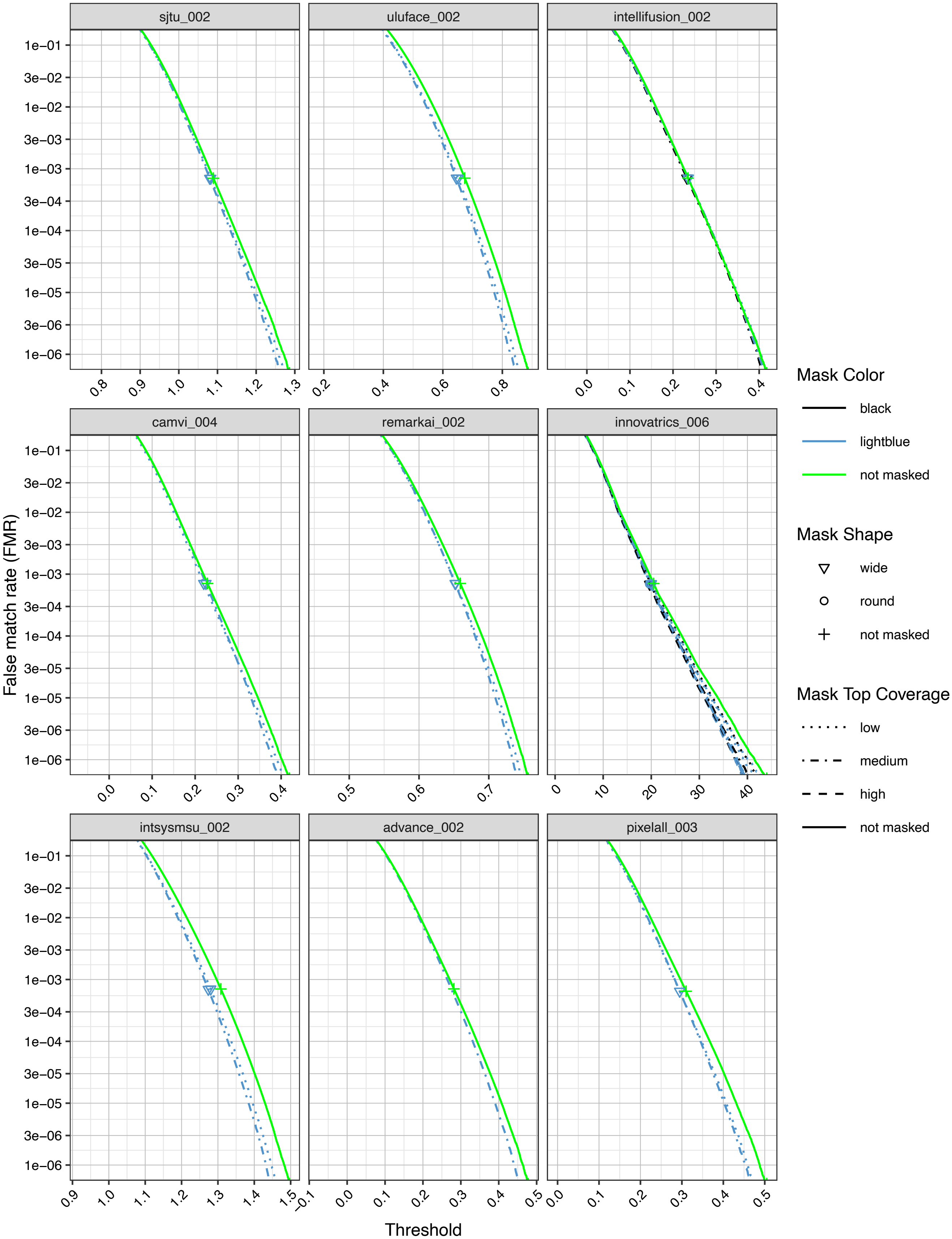


Figure 29: FMR calibration curves on unmasked and masked images.

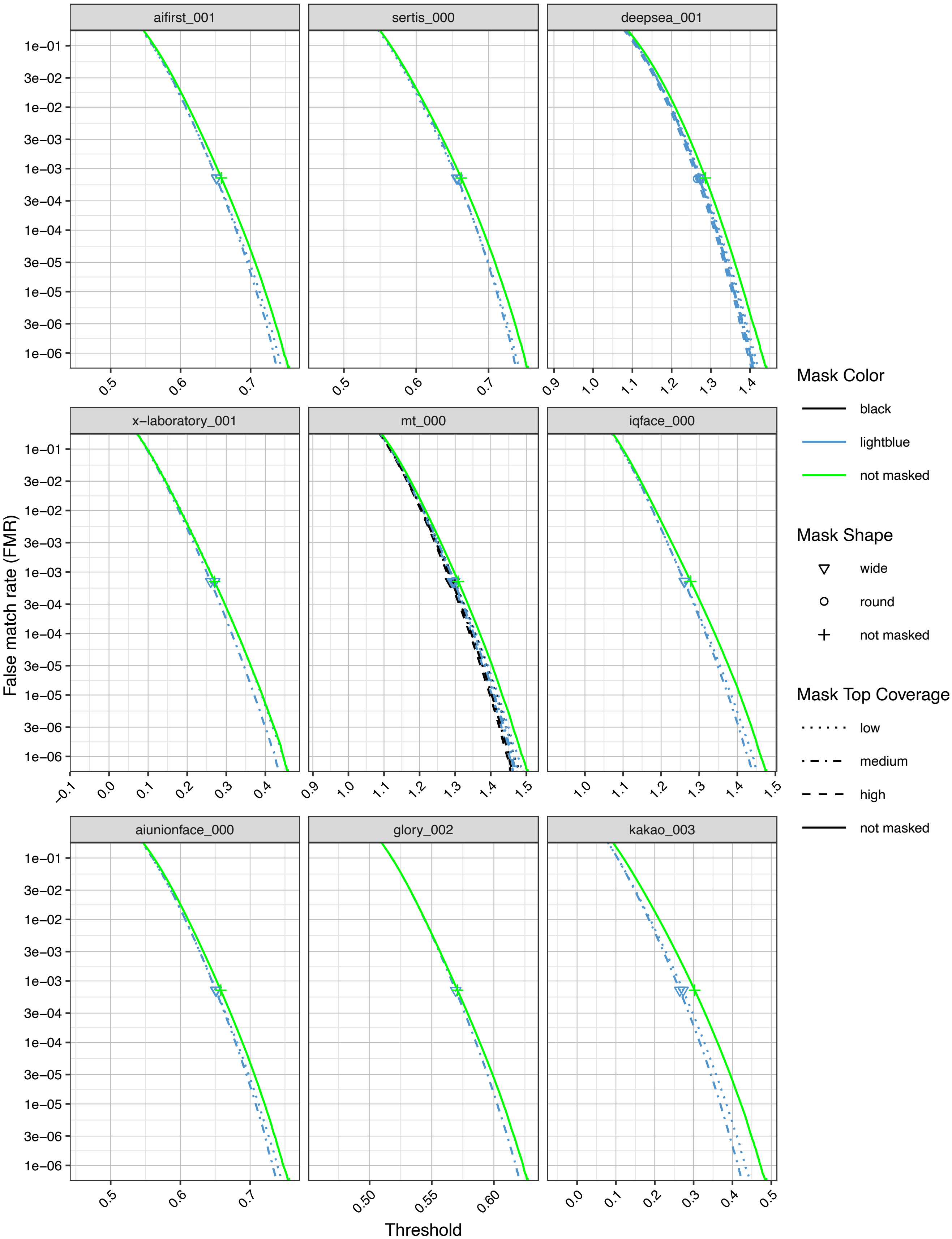


Figure 30: FMR calibration curves on unmasked and masked images.

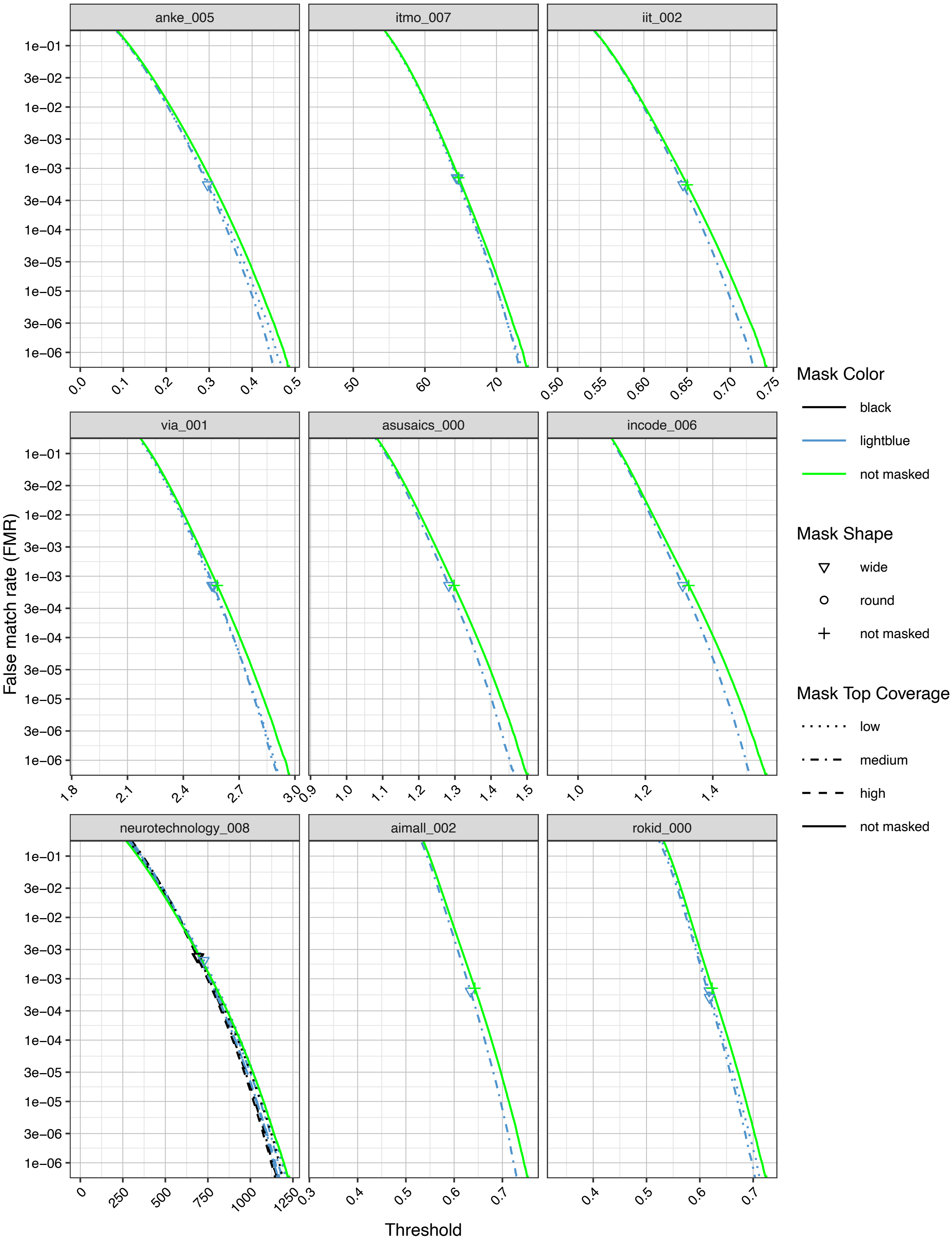


Figure 31: FMR calibration curves on unmasked and masked images.

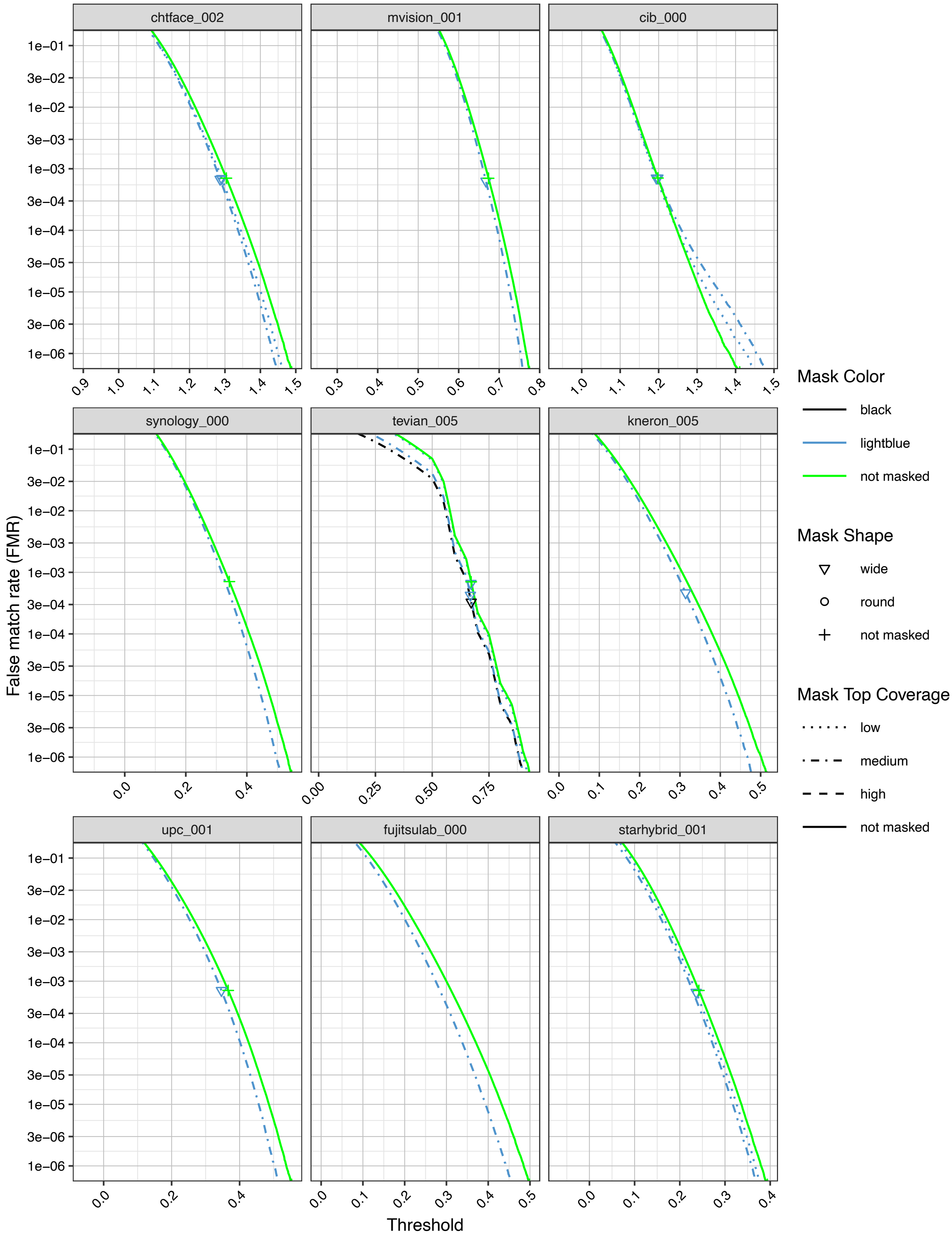


Figure 32: FMR calibration curves on unmasked and masked images.

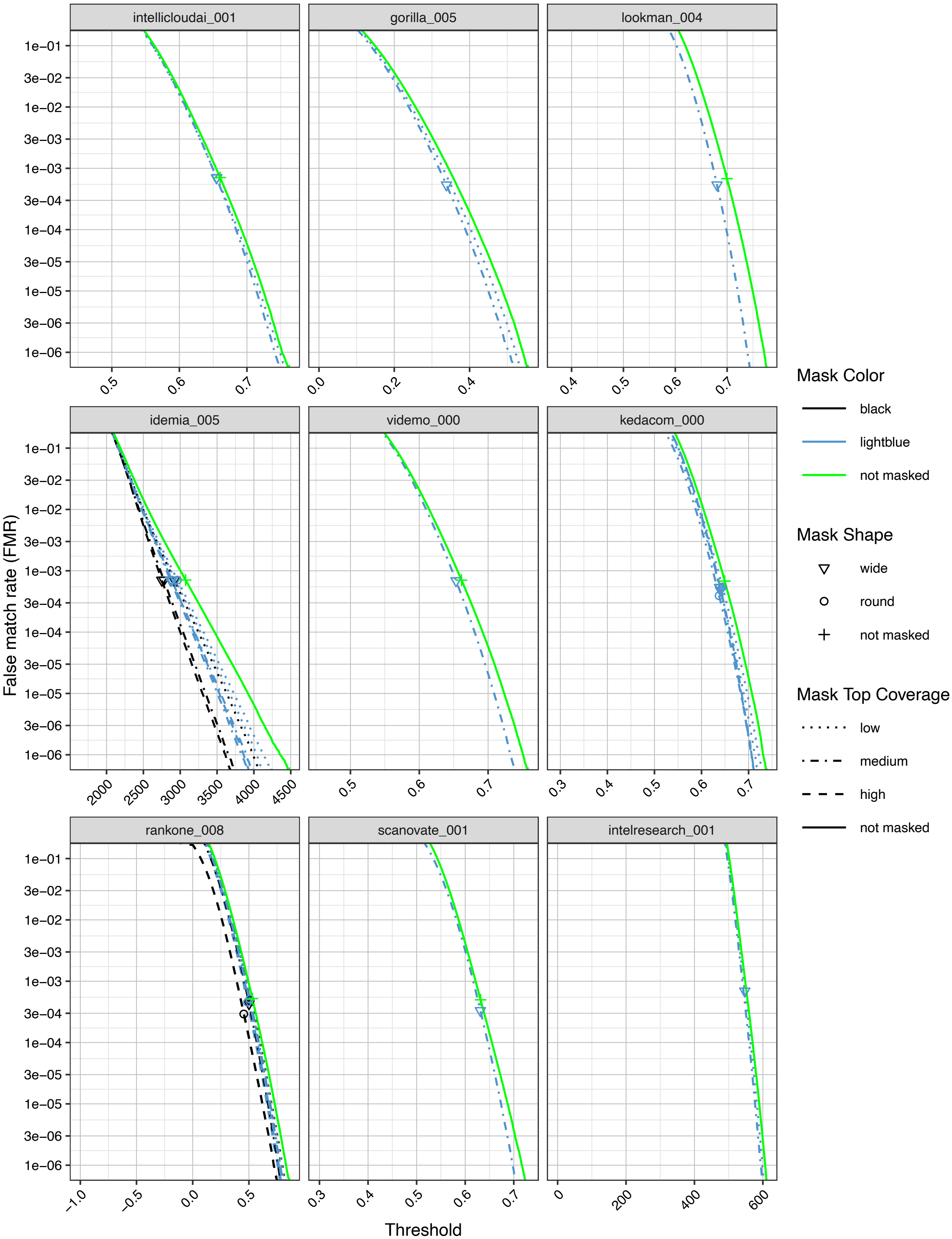


Figure 33: FMR calibration curves on unmasked and masked images.

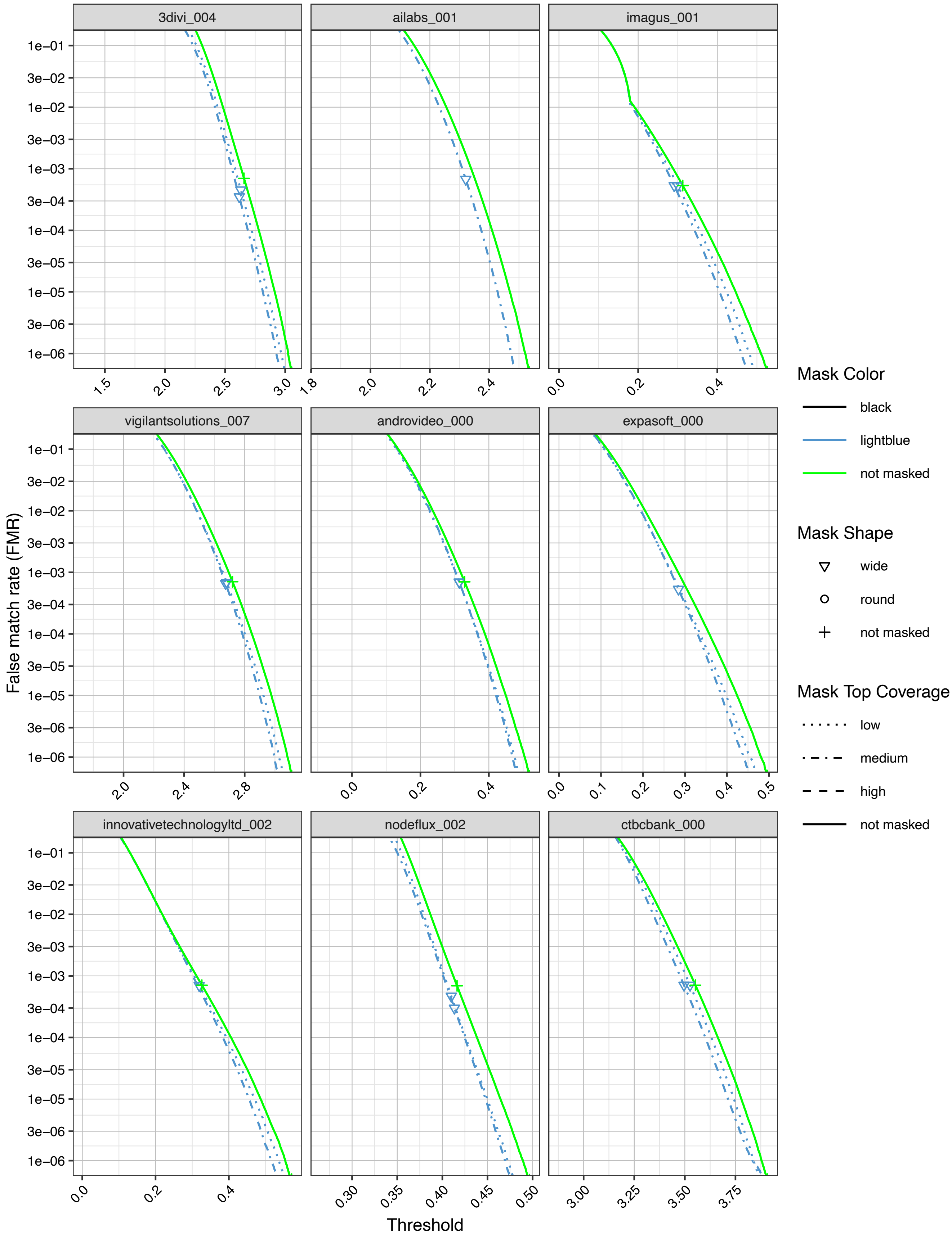


Figure 34: FMR calibration curves on unmasked and masked images.

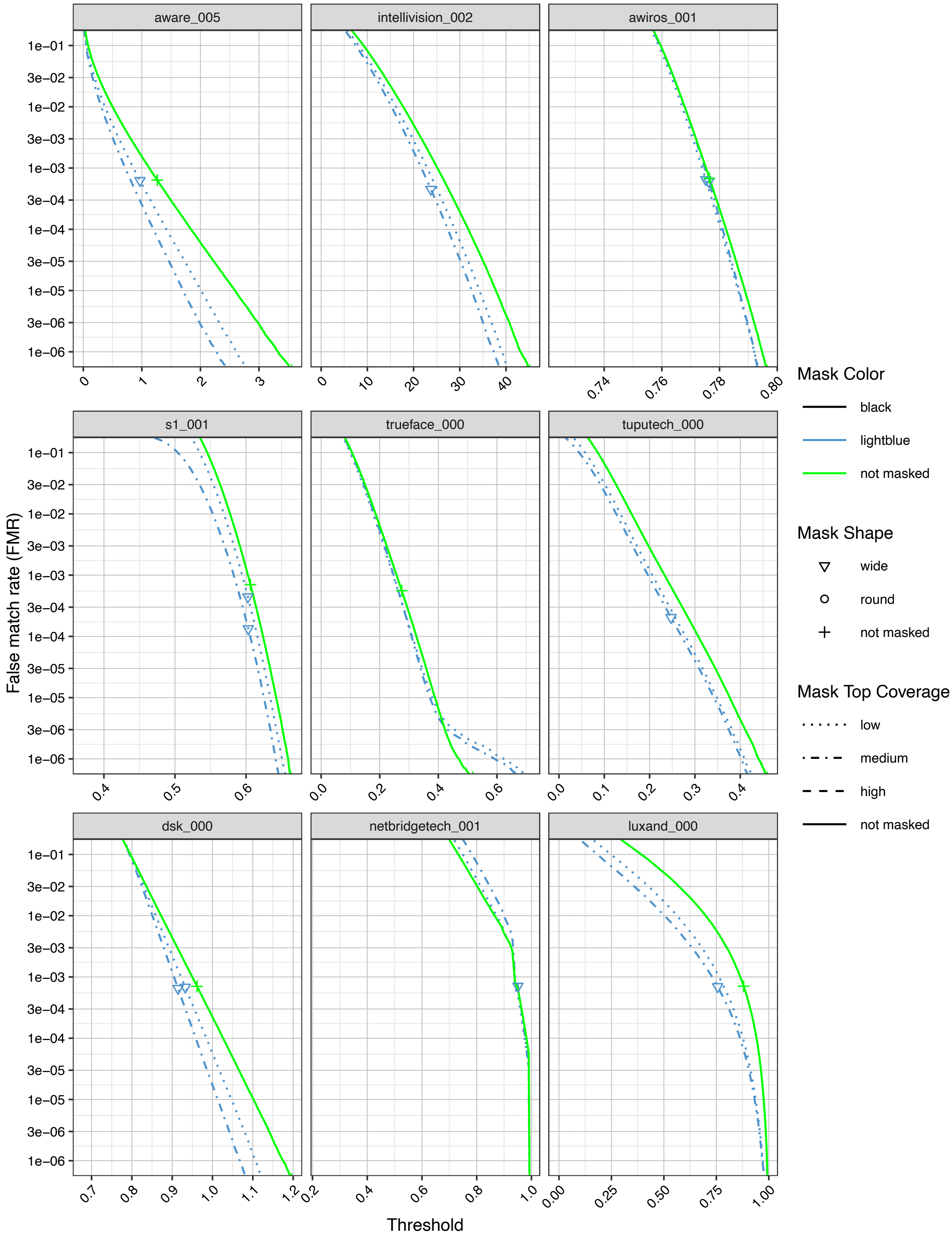


Figure 35: FMR calibration curves on unmasked and masked images.

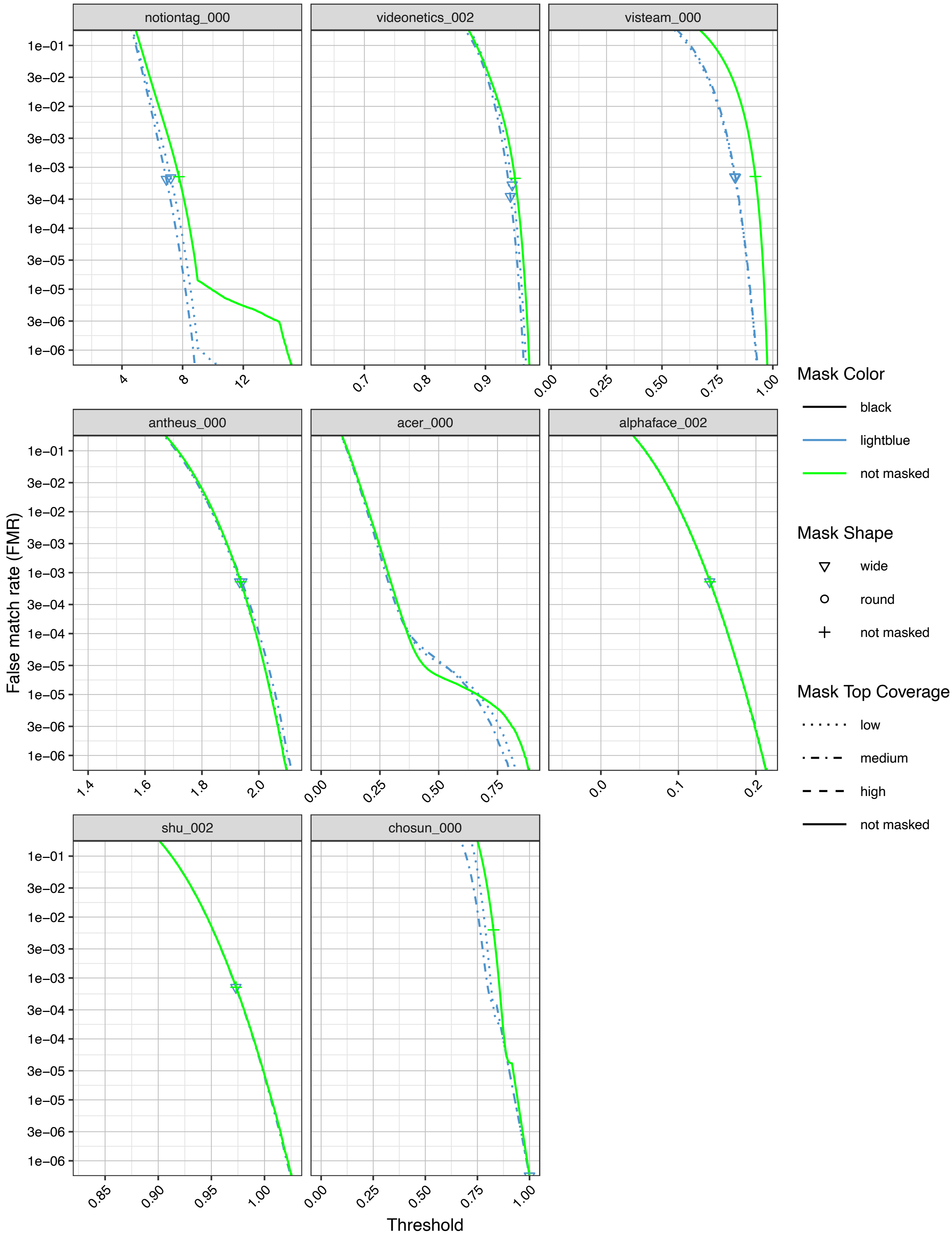


Figure 36: FMR calibration curves on unmasked and masked images.

References

[1] ISO/IEC 19794-5:2011 - Information technology - Biometric data interchange formats - Part 5: Face image data.

[2] N95 Respirators, Surgical Masks, and Face Masks. https://www.fda.gov/medical-devices/personal-protective-equipment-infection-control/n95-respirators-surgical-masks-and-face-masks.

[3] NIST Special Database 32 - Multiple Encounter Dataset (MEDS-II). https://www.nist.gov/itl/iad/image-group/special-database-32-multiple-encounter-dataset-meds.

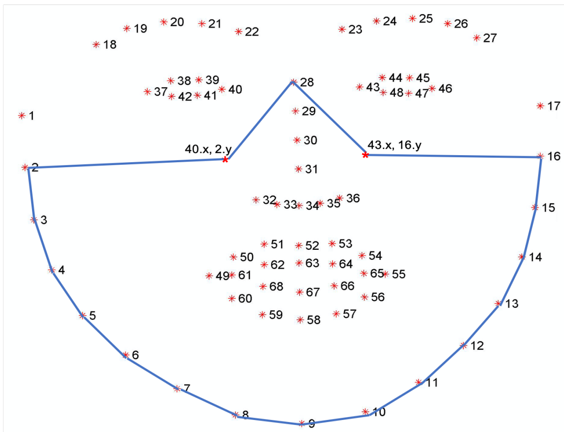
[4] Patrick Grother, George Quinn, and Mei Ngan. Face in video evaluation (five) face recognition of non-cooperative subjects. Interagency Report 8173, National Institute of Standards and Technology, March 2017. https://doi.org/10.6028/NIST.IR.8173.

[5] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009. http://dlib.net.

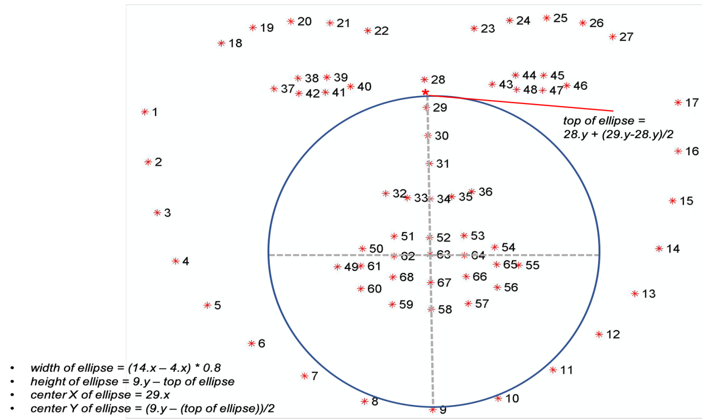
[6] Patrick Grother and Mei Ngan and Kayee Hanaoka. NIST Ongoing Face Recognition Vendor Test (FRVT) 1:1 Verification Application Programming Interface (API), April 2019. https://pages.nist.gov/frvt/api/FRVT_ongoing_11_api_4.0.pdf.

[7] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, Heling Chen, Yu Miao, Zhibing Huang, and Jinbi Liang. Masked face recognition dataset and application, 2020.

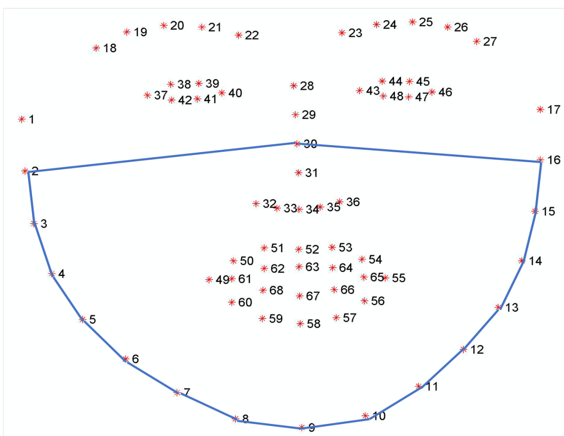
Appendix A Dlib Masking Methodology



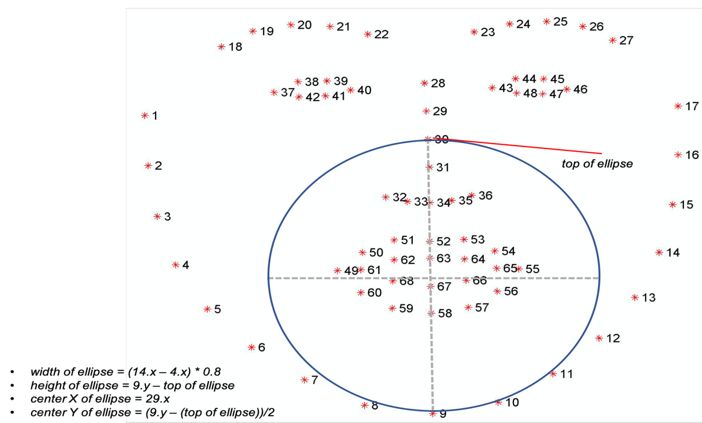
(a) wide, high coverage



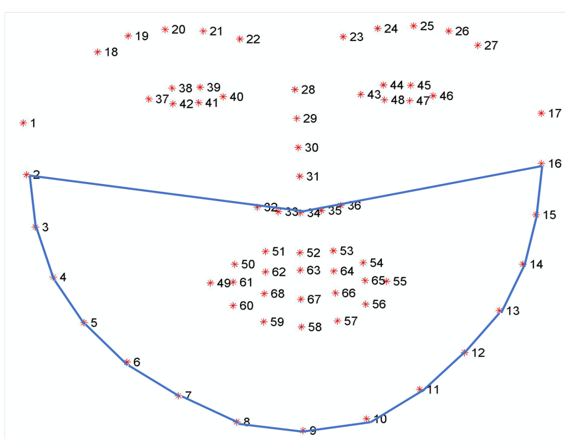
(b) round, high coverage



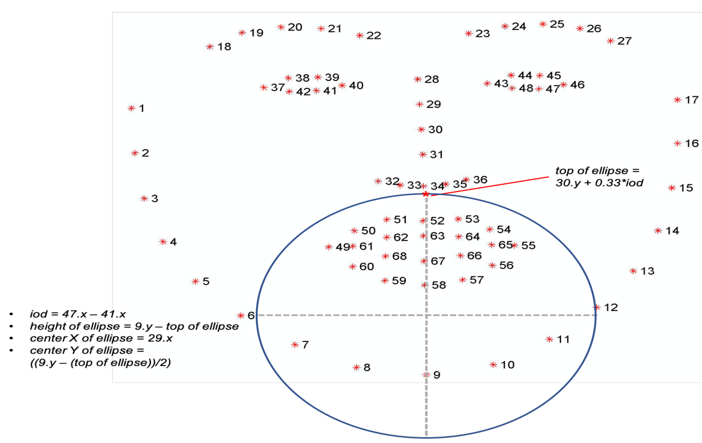
(c) wide, medium coverage



(d) round, medium coverage



(e) wide, low coverage



(f) round, low coverage

Figure 37: This figure shows the Dlib facial points used to create the various synthetic masks used in this report. For wide masks, the specified Dlib facial points were used to generate a closed polygon and two additional points were interpolated between each dlib facial point used for smoothing purposes. For round masks, the specified Dlib facial points were used to generate an ellipse. The Dlib C++ toolkit version 19.19, configured with the common histogram of gradients (HoG)-based face detector and 68 face landmark shape predictor was used to generate the 68 facial landmarks.