

隐私增强巩膜分割基准测试竞赛：SSBC 2025

M. Vitek^{1,*}, D. Tomašević^{1,*}, A. Das², S. Nathan⁴, G. Özbülak^{5,6}, G. A. T. Özbülak^{7,8}, J. P. Calbimonte^{8,9}, A. Anjos^{5,6}, H. H. Bhatt¹⁰, D. D. Premani¹⁰, J. Chaudhari¹⁰, C. Wang¹¹, J. Jiang¹², C. Zhang¹¹, Q. Zhang¹², I. I. Ganapathi¹³, S. S. A. D. Velayudan¹³, M. Assefa¹³, N. Werghi¹³, Z. A. Daniels¹⁴, L. John¹⁵, R. Vyas¹⁵, J. N. Khiarak¹⁶, T. A. Saeed¹⁷, M. Nasel¹⁸, A. Kianfar¹⁹, M. P. Panahi²⁰, G. Sharma²¹, P. R. Panth²¹, R. Ramachandra²², A. Nigam²¹, U. Pal³, P. Peer¹, V. Štruc¹

¹ University of Ljubljana (UL, SI), ² Birla Institute of Technology & Science Pilani (BITS Pilani, IN), ³ Indian Statistical Institute (ISI, IN), ⁴ Couger Inc., ⁵ École Polytechnique Fédérale de Lausanne (EPFL, CH), ⁶ Idiap Research Institute (CH), ⁷ Université de Lausanne (UNIL, CH), ⁸ University of Applied Sciences and Arts Western Switzerland (HES-SO, CH), ⁹ The Sense Innovation and Research Center (CH), ¹⁰ Ahmedabad University, ¹¹ Beijing University of Civil Engineering and Architecture (BUCEA, CN), ¹² People's Public Security University of China (PPSUC, CN), ¹³ Khalifa University of Science and Technology (KU, AE), ¹⁴ SRI International (US), ¹⁵ Pandit Deendayal Energy University (PDP, IN), ¹⁶ Warsaw University of Technology (WUT, PL), ¹⁷ Pirogov Russian National Research Medical University (RNRMU, RU), ¹⁸ Amirkabir University of Technology (AUT, IR), ¹⁹ Institute for Advanced Studies in Basic Sciences (IASBS, IR), ²⁰ Seraj Institute (SI, IR), ²¹ Indian Institute of Technology Mandi (IIT-M, IN), ²² Norwegian University of Science and Technology (NTNU, NO)

Abstract

这篇论文总结了 2025 年巩膜分割基准比赛 (SSBC)，该比赛聚焦于使用合成生成的眼部图像训练隐私保护的巩膜分割模型。比赛的目标是评估在合成数据上训练的模型相较于在现实世界数据集上训练的模型的表现。比赛共设有两个赛道：(i) 仅依赖于合成数据进行模型开发，以及 (ii) 将合成数据与 (少量的) 真实世界数据混合/结合进行模型开发。共有九个研究小组提交了多样化的分割模型，采用了一系列的架构设计，包括基于 transformer 的解决方案、轻量级模型和由生成框架指导的分割网络。实验在三个评估数据集上开展，这些数据集包含在各种条件下收集的合成和真实世界图像。结果显示，完全在合成数据上训练的模型可以获得有竞争力的表现，尤其是在应用专门的训练策略时，如在合成数据赛道中表现最佳的模型取得了超过 0.8 的 F_1 得分。此外，在混合赛道中的性能提升通常更多是由方法选择驱动，而不是因为包括了真实数据，这体现了合成数据在隐私意识的生物识别开发中的潜力。比赛的代码和数据可在以下地址获得：https://github.com/dariant/SSBC_2025。

1. 介绍

眼睛生物特征代表了一个热门的研究分支，该分支专注于计算机辅助技术，能够根据独特的眼睛特征推断个人身份。尽管虹膜识别在历史上主导了这一领域，最近的研究越来越关注其他可以补充或替代虹膜的眼睛特征。在这些特征中，眼白（即眼睛的白色区域）由于其独特性、长期一致性和对呈现攻击的抵抗力 [12, 46, 59]，而成为一个有前景的候选项。与虹膜不同，眼白不易受到常见视觉状况的影响，可以在正常光线下用标准相机拍摄，这使其在实际应用中尤为实用。

*M. Vitek and D. Tomašević are first authors with equal contributions.

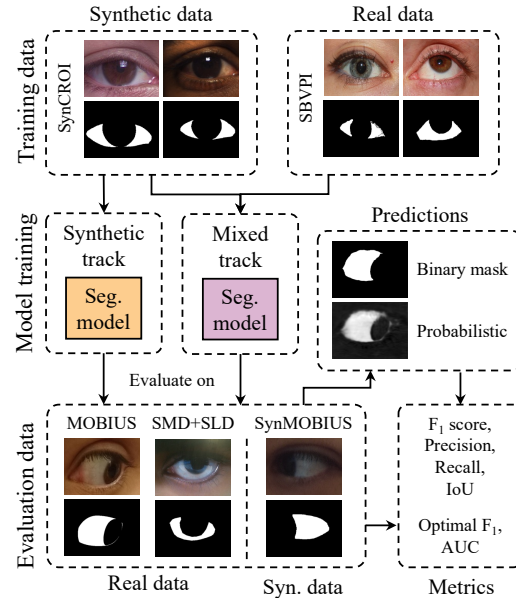


Figure 1. 2025 年 SSBC 概览。比赛包含两个赛道，其中巩膜分割模型在 (i) 合成数据和 (ii) 合成和真实数据的混合上进行了训练。模型在隔离的真实和合成数据集上进行评估。参与者需要提交二值和概率预测。

近年来，巩膜生物特征研究趋于活跃，涉及多种主题，包括识别算法 [2, 10, 15, 20, 21, 43, 59, 60, 65]、分割方法 [1, 41, 46, 56, 57]、展示攻击检测 [15, 46, 59]、用户变异性的适应性 [8, 11]、融合技术 [17, 26] 和轻量级方法 [25, 56, 60]。然而，由于该领域的研究总体上正在向数据需求量大的深度学习模型发展，一个出现的关键问题是隐私保护，因为考虑到隐私保护措施而编制的大规模眼部图像数据集在很大程度上是不可用的。解决这一问题的一个可能方法是使用合成生成的数据 [9, 51–54]，这些数据保留了开发生物特征模型所需的特征，但不属于（或不包含有关）真实个人的识别

信息，因此不存在被滥用或隐私泄露的风险。

为了研究合成数据在巩膜生物识别中的有效性, 2025 年巩膜分割基准测试竞赛 (SSBC), 如 Fig. 1 中总结的, 在国际生物识别联合会议 (IJCB 2025) 的框架内组织。该竞赛专注于巩膜分割, 这是任何巩膜识别流程中的关键步骤, 因为分割质量直接影响所有后续阶段, 包括归一化、特征提取和身份匹配。竞赛旨在回答关键的研究问题, 例如: 合成数据与真实世界数据集的相似程度如何? 合成训练数据的使用对分割性能有什么影响? 在训练过程中, 哪些技术可以用于使模型适应合成生成的图像? 来自世界各地的九个可见研究团队提交的分割模型被评估, 以提供对这些及相关问题的见解。主办方与参赛团队的共同努力带来了以下贡献:

- 对真实世界和合成生成的评估数据中的当代巩膜分割模型进行全面评估。
- 一个差异化性能分析, 研究合成和真实世界训练数据在模型开发中的影响。
- 研究模型大小和计算复杂度对最终分割性能及模型对合成训练数据适应性的影响。

2. 相关工作

SSBC 2025 是一项巩膜分割基准竞赛的第 9th 次迭代, 最初于 2015 年在 BTAS 会议上开始。SSBC 系列竞赛显著推动了巩膜分割模型的发展, 每次迭代都解决了不同的研究问题。1st 和 3rd SSBC (SSBC 2015 和 SSBC 2016) 研究了各种模型的分割性能, 并另外引入了新的巩膜分割数据集 (即 MASD 和 SMD) [14]。2nd 次迭代 (SSRBC 2016) 除了研究巩膜分割技术外, 还研究了识别方法 [16]。4th 迭代, SSRBC 2017, 再次包括识别任务, 但还探讨了视线方向对成功分割和识别的影响 [18]。5th 竞赛, SSBC 2018, 研究了跨传感器图像捕获对巩膜分割性能的影响 [19], 而 6th 次迭代, SSBC 2019, 调查了跨分辨率环境如何影响分割性能 [13]。7th 版, SSBC 2020, 引入了一个专门为移动巩膜生物特征编制的新数据集 (MOBIUS), 因此专注于移动领域的巩膜分割 [58]。SSBC 2020 的后续工作 [57] 随后探讨了当代巩膜分割模型中存在的各种偏差。最后, 8th 次迭代 (SSRBC 2023) 分别研究了分割和识别性能, 以及两项任务之间的相互作用 [7]。

与以往的 SSBC 版本不同, SSBC 2025 重点关注隐私保护的巩膜分割模型, 这些模型是利用合成生成的 (无身份) 训练数据开发的。比赛的主要 (不同) 目标是研究这些模型在性能上与基于真实个人数据训练的模型相比表现如何。

3. SSBC 2025 竞赛数据

SSBC 的目的是研究使用合成数据进行巩膜分割模型的训练和测试, 并调查与现实世界巩膜图像相比, 合成数据对模型性能的影响。为此, 竞赛中使用了若干现实世界 (R) 和合成数据集 (S), 即: (i) 巩膜血管、眼周和虹膜 (SBVPI) 数据集 (R) [59], (ii) 非受限环境中的移动眼部生物识别技术 (MOBIUS) 数据

Table 1. 用于 SSBC 2025 的真实和合成数据集的总结。报告包括图像和被试的数量、主要变异来源及其在竞赛中的目的。

Origin	Dataset	# Images	# IDs	Variability	Purpose
Real-world	SBVPI	1840	55	GZ, CLR	Training
	SMD+SLD	489	52	CN	Testing
	MOBIUS	3540	35	GZ, CLR, CN	Testing
Synthetic	SynCROI (CE)	5500	N/A	GZ, CLR	Training
	SynCROI (PU)	5500	N/A	CN	Training
	SynMOBIUS	4772	N/A	GZ, CLR, CN	Testing

(GZ) - gaze, (CLR) - eye color, (CN) - acquisition condition

集 (R) [58], (iii) 组合的巩膜移动数据集 [6] 和巩膜活体检测数据集 (R) [7] (即 SMD+SLD), 以及 (iv) 合成跨种族眼部图像 (SynCROI) 数据集 (S), 和 (v) 合成 MOBIUS (SynMOBIUS) 数据集 (S)。SBVPI 和 SynCROI 是 SSBC 2025 所用的主要训练数据集, 而 MOBIUS、SMD+SLD 和 SynMOBIUS 仅用于评估目的。数据集的总结在表 1 中提供, 每个数据集的样本图像及其对应的分割掩码在图 2 中展示。作为 SSBC 2025 的一部分, 所有数据集的图像被调整为 400 × 300 的分辨率。

3.1. 真实世界数据集

SBVPI。SBVPI 是比赛中的第一个真实世界数据集, 由 1840 张高分辨率 RGB 图像组成, 这些图像来自 55 名高加索被试者 (29 名女性和 26 名男性), 使用佳能 EOS 60D 单反相机及微距镜头在受控条件下拍摄。图像包含每只眼睛的四个注视方向 (直视、左视、右视和上视)。在不同距离和相机位置下获取每只眼睛和注视方向的 4 个样本。被试者的年龄范围是 15 到 80 岁, 眼睛颜色多样。所有图像都有手动标注的巩膜和眼周区域分割掩码, 且对于 100 张图像, 还有更为详细的掩码, 包括虹膜和瞳孔。

SMD+SLD。第二个真实世界数据集, SMD+SLD, 将巩膜移动数据集 [6] 和巩膜活体数据集 (SLD) [57] 结合在一起, 包含来自 SMD 的 25 名受试者的 381 张 RGB 图像以及来自 SLD 的 27 名受试者的 108 张图像。这些图像是通过不同手机的 800 万像素后置摄像头在各种采集条件下捕获的, 以增加数据多样性, 导致图片模糊、眼睛眨动的图像以及在一天中不同时间和不同光照条件下拍摄的图像。结果, 该数据集能够评估巩膜分割模型在非理想或具有挑战性的条件下的表现。数据集还包含为巩膜区域手动生成的真实分割掩码。

MOBIUS。比赛的第三个真实世界数据集 MOBIUS, 是专门针对移动眼球生物识别设计的。整个数据集包含来自 100 名高加索裔男性和女性受试者的 16,717 张高分辨率 RGB 双眼图像。使用三种商用移动设备 (Sony Xperia Z5 Compact, Apple iPhone 6s 和小米 Pocophone F1) 在不同的注视方向 (直视、左视、右视、上视) 和光照条件 (自然日光、室内光、低光) 下拍摄, 从而带来了多样性高的图像。对于 SSBC 2025 而言, 仅使用了包含手动标注分割掩码的数据集部分, 包括来自 35 名受试者的 3542 张图像, 并标注了巩膜、虹膜和瞳孔区域。然而, 在 SSBC 2025 的评估过程中,

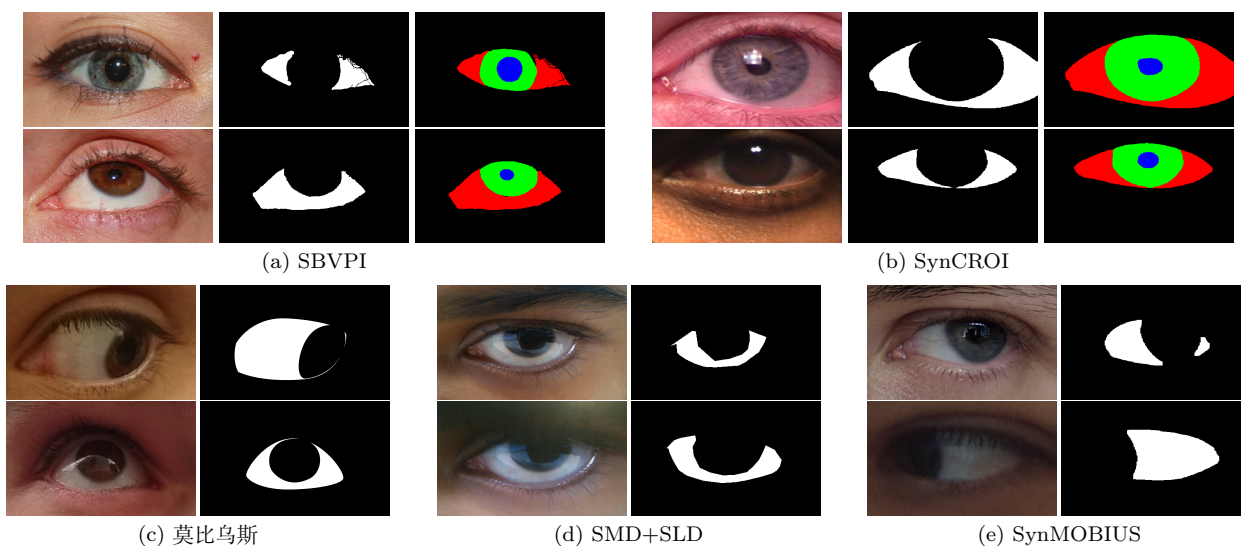


Figure 2. 来自 SSBC 2025 的训练和测试数据集的样本。上排包含训练数据集的眼部图像和可用的真实分割掩码，而下排展示测试样本及其参考掩码。SSBC 2025 的参与者可以自由选择完整的四类标记上训练他们的模型，或者仅在对应于巩膜的区域进行训练。

仅考虑了巩膜区域。

3.2. 合成数据集

由于 SSBC 2025 专注于开发使用隐私保护合成数据的分割模型，我们构建了两个大规模合成数据集，SynCROI 和 SynMOBIUS，用于训练和评估，遵循如图 3 所述的方法。

合成数据生成。为生成合成数据集，我们依赖于 BiOcularGAN [53]，这是一种最新的深度生成框架，用于创建具有对应（合成、真实）分割掩码的双模态眼部图像。BiOcularGAN [53] 扩展了 StyleGAN2 [33] 方法并通过使用双分支合成网络来实现更好的图像质量，该网络基于映射网络提供的噪声驱动的样式信息创建对齐的可见和近红外（NIR）眼部图像，同时具有两个鉴别器网络，每个光谱一个，形成一个联合训练目标。此外，BiOcularGAN 依赖于一个独立的像素分类器集成来解释合成网络的潜在特征，以准确生成合成图像的掩码 [64]。关键的是，该模型还采用了自适应鉴别器增强（ADA），即使在低数据环境中也能实现稳定训练 [32]。我们仅使用生成的 RGB 图像用于比赛，放弃生成的 NIR 图像。

我们采用的具体 BiOcularGAN [53] 实现包括一个 8 层的映射网络，一个具有 7 个合成块的合成网络，输出分辨率从 4×4 到 256×256 的 VIS-NIR 数据，以及两个每个具有 7 个下采样块的判别器。我们以 16 的批量大小训练模型，学习率为 0.0025，使用 Adam 优化器 [34]，超参数为 $\beta_1 = 0$ 、 $\beta_2 = 0.99$ 和 $\epsilon = 10^{-8}$ 。用于生成合成数据的真实标签的集成像素分类器，由 10 个多层感知器组成，这些感知器在初始生成模型后训练，用于随机采样的手动标注合成图像的像素。训练使用交叉熵损失和 Adam 优化器 [34]，在 64 的批量大小下，学习率为 10^{-3} ，训练在第三个迭代周期后如

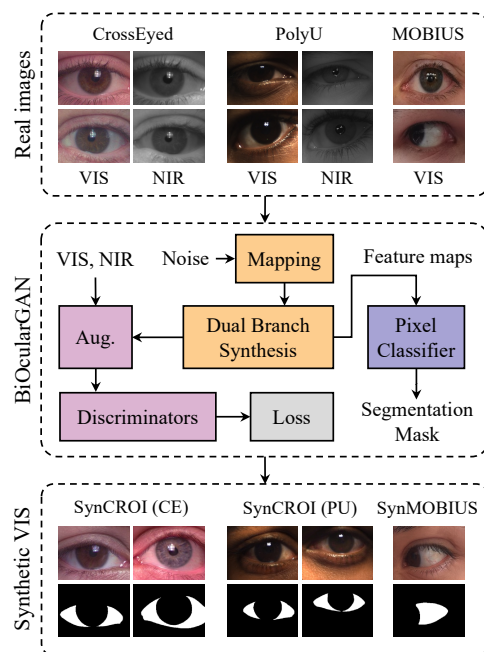


Figure 3. 数据合成过程概述。为了生成 SSBC 2025 的合成数据集，我们分别在三个真实世界的数据集（Cross-Eyed [49]，PolyU Cross-Spectral Iris，和 MOBIUS [58]）上训练 BiOcularGAN [53] 的独立实例。

果无改进则停止于 50 步后。使用 BiOcularGAN [53]，我们为 SSBC 2025 生成以下合成数据集：

- SynCROI。第一个合成数据集，SynCROI，包含 11,000 张合成眼部图像，每张图像都配有巩膜、虹膜和瞳孔区域的分割掩码。该数据集分为两个大型子集，包含不同来源的主体，每个子集都通过分别训练于真实可见光（VIS）和近红外（NIR）图像

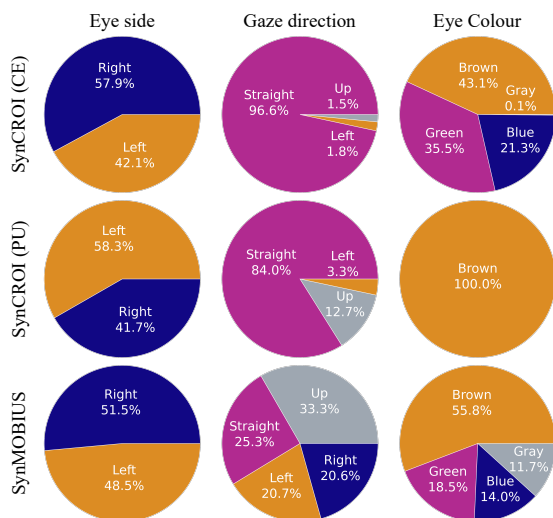


Figure 4. 合成样本在不同特征上的分布：眼睛侧别（左、右）、注视方向（正前、左、右、上）和眼睛颜色（棕色、蓝色、灰色、绿色）。这些值是通过在 MOBIUS 和 SBVPI 注释上训练的基于 ResNet101 的 [28] 分类器估算的。

对的 BiOcularGAN [53] 实例生成。第一个子集，SynCROI (CE)，是通过在 CrossEyed (CE) 数据集上训练 BiOcularGAN [53] 生成的，该数据集包括 3840 对具有不同眼色的 120 白人主题 VIS-NIR 图像。不同的是，第二个子集，SynCROI (PU)，基于香港理工大学 (PolyU) 跨光谱虹膜图像数据库，该数据库有 12,540 对亚洲主体的 209 VIS-NIR 图像。这些模型分别在 1120 和 1600 千张图像上训练到收敛。一旦训练完成，每个模型生成 8 个样本，并手动标注用于像素分类器的训练。之后，随机对噪声输入进行采样以便每个模型产生 5500 张对齐的 VIS 和 NIR 图像，这些图像为 256×256 像素，并配有相应的分割掩码。在比赛中，仅考虑 VIS 图像及其掩码，并调整到 400×300 以保持真实世界训练数据的比例。

- SynMOBIUS。第二个合成数据集，SynMOBIUS，源自 MOBIUS [58] 数据集，专门用于 SSBC 2025 的评估。与之前的设置不同，首先会调整 BiOcularGAN [53] 模型，通过移除第二个合成分支及其相关判别器来适应单一 (VIS) 光谱图像的训练。经修改的模型在 MOBIUS [58] 上进行了 2240 千幅图像的收敛性训练。为了训练像素分类器，我们利用来自 MOBIUS [58] 的 45 手动标注，通过将其对应的图像投射到生成模型 [33] 的潜在空间中，以及包含眼部遮挡的 5 手动标注的合成样本。在训练后，该模型被用来生成 5000 图像-掩码对，随后将其从 256×256 调整到 400×300 。

合成数据的特征。为了更好地了解生成的合成数据集，我们利用基于 ResNet101 的 [28] 分类器来确定每个合成图像的视觉特征，包括眼睛的方向（左或右）、注视方向（直视、左、右或上）以及眼睛颜色（棕色、蓝色、灰色或绿色）。对于 SynCROI 数据集，我们在包含必



Figure 5. 待提交结果的说明（从左到右）：原始图像，生成的二值分割掩码，概率（灰度）分割预测。

要注解的 SBVPI 和 MOBIUS 的组合数据上训练了分类器。训练是在 9 : 1 数据分割上进行的，跨越 20 个周期，以 32 为批量大小，并使用 Adam 优化器 [34]，初始学习率为 10^{-4} ，如果在 4 个周期中没有观察到改善，则学习率按 10 的因子减少。相反，为了最小化领域差距，SynMOBIUS 的分类器仅在 MOBIUS 数据集上进行训练，参数相同。每个数据集预测的特征分布如图 4 所示。注意，SynCROI 的两个子集主要包含正视注视方向的图像，并且由于生成模型训练数据中这些特征的普遍存在，SynCROI (PU) 仅包含棕色眼睛。

4. 基准测试方法

4.1. 竞赛协议

SSBC 2025 包含两个独立的阶段。第一阶段中，参与者被提供了来自 SBVPI 和 SynCROI 数据集的现实世界和合成训练数据，这些数据以眼部图像和真实分割掩码的形式提供。数据集被分为训练和验证子集，但参与者不需要使用预先分配的分割。为了让参与者熟悉最终提交格式，还发布了一小部分测试数据，包括来自 MOBIUS 和 SMD+SLD 的单一身份图像，以及 10 个来自 SynMOBIUS 的随机样本。这些隔离的测试数据在比赛开始 9 周后提供给参与者，并未提供真实分割掩码。然后，参与者只有一周时间来提交他们的分割预测，以限制可能在测试数据上的微调。参与者需要为来自 MOBIUS、SMD+SLD 和 SynMOBIUS 的每个 14,772 图像提供两种类型的输出，包括：(i) 二值分割掩码，其中非零像素表示巩膜区域，零值像素表示背景，以及 (ii) 灰度概率图，像素强度表示该像素属于巩膜区域的置信度。预期输出格式的示例见图 5。提交的二值掩码作为参与模型性能排名的基础，而概率图则用于获得对模型行为的更深入理解，并使得详细性能曲线的推导成为可能。

评估是在两个不同的轨道中进行的，这反映了 SSBC 2025 在模型开发和评估中关于使用合成生物特征数据保护隐私的重点。第一，(i) 合成轨道专注于仅在 SynCROI 数据集的合成数据上训练的模型，研究这些模型在不同的、被隔离的合成和真实世界评估数据上的表现。第二，(ii) 混合轨道则允许参与者在合成 SynCROI 数据和真实世界 SBVPI 数据的混合上训练模型，并且具体比例由他们自行决定。与以前一样，训练的模型在合成和真实世界的评估数据上进行评估。

4.2. 性能指标

分割模型的整体性能通过提交的二元分割掩码和概率预测进行评估。对于二元掩码，标准的分割指标是由

Table 2. SSBC 2025 提交条目的列表及其所属机构。本表中使用的模型缩写与实验部分中使用的缩写相对应。

Team	Model Acronym	Base [†]
Indian Institute of Technology Mandi (IIT-M)	SEG-U-Sclera	SAM
Beijing University of Civil Engineering and Architecture (BUCEA)	SAM2-UNet	SAM/UN
Warsaw University of Technology (WUT)	AEOS	SAM/SF
Khalifa University (KU)	KU-CVML	DL
Ahmedabad University (AU)	ShapeGAN-DLV3+	DL
Couger Inc.	SwinDANet	ST/DN
SRI International	UL-VMUNet	M/UN
Idiap – University of Applied Sciences and Arts Western Switzerland (HES-SO)	SAM-Iris	SAM
Pandit Deendayal Energy University (PDPU)	UNet++_Binary	UN

For details on the participants from the institutions, see the author list.

[†] Base architecture: DL – DeepLab [5], DN – DenseNet [30], M – Mamba [27], SAM – Segment Anything Model [35], SF – SegFormer [62], ST – Swin Transformer [37], UN – U-Net [44].

真正例 TP （即正确检测到的巩膜像素）、假正例 FP （即背景像素错误检测为巩膜）和假负例 FN （即巩膜像素错误检测为背景）导出的，如下所示：

- 精确度衡量的是正确预测为巩膜的像素相对于被预测为巩膜的像素总数的比例，即 $\frac{TP}{TP+FP}$ [38,39,45]。
- 召回率量化了模型成功识别的真实巩膜像素的比例，计算方法为 $\frac{TP}{TP+FN}$ [23,38,39,45]。
- F_1 分数，表示精确率和召回率的调和平均数，定义为 $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ 。该聚合指标平衡了精确率和召回率之间的权衡，并作为排名参与模型的主要标准。
- 交并比 (IoU)，或称 Jaccard 指数，抓取预测的巩膜区域和真实的巩膜区域之间的重叠部分，由它们的并集归一化，计算为 $\frac{TP}{TP+FP+FN}$ 。

为了提供比二元分类更全面的评估，还使用灰度（概率）分割图来生成精确召回曲线 [40,48]。从这些曲线中，可以计算出最佳的 F_1 分数 (F_1^{opt}) 和曲线下面积 (AUC) [4]，从而在不同的置信度阈值下对模型进行更细致的比较。

5. 提交方法的总结

总共有 9 个团队提交了他们的作品到 SSBC 2025。Table 2 展示了提交作品的总结，下面提供了对每个作品的简要描述。

SEG-U-Sclera (IIT-M) 是一种经过不确定性加权的二元交叉熵 (BCE) 训练的 SAM2 [42] 的变体。该损失通过关注合成和真实世界图像差异的区域来解决真实世界和合成数据之间的差异。

SAM2-UNet (BUCEA) 是另一种基于 SAM2 的 [42] 模型，采用类似 U-Net 的架构 [63]，包含一个 Hiera 编码器 [47] 和一个 U 形的 3 块解码器。在参数冻结的 Hiera 主干中插入轻量级适配器，以确保对巩膜分割任务进行参数高效的微调。该模型通过加权 IoU 和 BCE 损失以及深度监督策略进行训练。

AEOS (WUT)：WUT 团队采用了一种混合方法，他们的架构取决于训练数据的量。他们基于 SAM2 的 [42] 架构在训练数据量较少的情况下表现不佳，因此在合成赛道中，他们依赖于类似 SegFormer 的 [62] 模型。这

种混合方法提高了泛化能力，因为它明确处理了小/大训练数据集。

KU-CVML (KU) 依赖于 DeepLabV3+ [5] 架构和 EfficientNet-B4 [50] 编码器，并使用一种混合损失进行训练，该混合损失部分地受到半监督训练方法和 SAM [35] 的启发，同时也解决了类别不平衡和轮廓精度的问题。

ShapeGAN-DLV3+ (AU) 是 DeepLabV3+ [5] 的扩展，包含几个特定于巩膜分割的模块，即：(i) 结合了 Co-ordConv [36] 和可变形的干状卷积 [67]，以提供空间感知和自适应感受野；(ii) 一个 FourierLoss [24]，以确保形状的保真度；(iii) 一个 InverseFormLoss [3]，以促进精细边界对齐；以及 (iv) 一个 PatchGAN 识别器，连接到编码器的末端并以对抗方式训练，以跨越合成与真实领域之间的差距。

SwinDANet (Couger) 是一个混合编码器-解码器模型。该编码器基于带有平移窗口自注意力的 Swin Transformer V2 [37]，这使得能够对长程依赖关系进行建模，并在分层上下文表示中提取多尺度特征。解码器基于 DenseNet [30]，并通过 CSSE 注意力 [31] 增强，强调显著的空间和通道信息，以及自适应跳跃连接以保留来自编码器的空间细节。

UL-VMUNet (SRI) 是一种轻量级 Mamba [27] 变体，结合了状态空间 Mamba 模型和类似 U-Net 的架构 [61]，以实现计算和参数高效的高质量分割。为了更好地从合成数据中泛化，进行手工设计的数据增强以模拟不同的光照、姿势和相机伪影。

SAM-Iris (Idiap – HES-SO) 是一个基于 SAM2 的 [42] 模型，通过受 [55] 启发的顺序合成到真实训练过程进行微调。该模型额外包含一个在输入眼睛图像中心的提示点，确保模型专注于眼睛区域。

UNet++_Binary (PDPU) 采用 UNet++ [66] 架构，使用 ResNet 风格的 [29] 编码器。它针对使用 ImageNet [22] 统计数据归一化的图像上的巩膜分割任务进行了微调。

6. 基准测试结果

在本节中，我们展示了 SSBC 2025 的结果，涵盖了合成轨道和混合轨道。此外，我们对不同分割阈值下提交模型的性能进行了详细分析，并研究了来自两个比赛轨道的相应模型之间的性能差异。

6.1. 总体结果和性能排名

为了评估和排名 SSBC 2025 的提交结果，我们计算了提交的分割掩码的平均性能分数。本节中报告的标准误差是通过将测试数据分成 5 个不重叠的主体数据集并计算相应的标准偏差得到的。跨三个评价数据集计算的调和平均 F_1 分数被用作模型排名的标准。这确保了评分系统奖励在所有评价数据集中保持稳定高性能的提交，而不是在某些数据集上表现良好而在其他数据集上失败。

Table 3. 对基于合成数据训练的模型进行比较评估。提交的结果根据三个评估数据集上获得的 F_1 分数的调和平均值进行排名。 F_1 、精确度、召回率和 IoU 分数是根据提交的二进制掩码计算的。精确度-召回率曲线上的最佳 F_1 分数 (F_1^{opt}) 和 AUC 值是从概率分割预测中计算得出的。在紧邻各个数据集结果的列中还报告了其他性能指标的调和平均值。需要注意的是，排名在各个性能指标上是相当一致的。

Rank	Segmentation Model	Evaluation Dataset	From binary masks							From probabilistic predictions			
			F_1		Precision		Recall		IoU		F_1^{opt}		AUC
1	SwinDANet	MOBIUS	0.824		0.825		0.842		0.722		0.850	0.824	0.887
		SMD+SLD	0.725	0.798	0.640	0.754	0.866	0.868	0.587	0.680	0.758	0.809	0.870
		SynMOBIUS	0.856		0.830		0.897		0.757		0.875		0.921
2	SAM2-UNet	MOBIUS	0.792		0.792		0.822		0.677		0.821		0.843
		SMD+SLD	0.729	0.776	0.636	0.728	0.872	0.856	0.584	0.649	0.756	0.803	0.779
		SynMOBIUS	0.813		0.777		0.876		0.698		0.837		0.861
3	KU-CVML	MOBIUS	0.783		0.711		0.914		0.667		-	-	-
		SMD+SLD	0.717	0.770	0.596	0.676	0.926	0.925	0.571	0.641	-	-	-
		SynMOBIUS	0.816		0.738		0.937		0.700		-	-	-
4	UL-VMUNet	MOBIUS	0.755		0.822		0.736		0.633		0.817		0.859
		SMD+SLD	0.689	0.744	0.600	0.728	0.839	0.793	0.537	0.609	0.729	0.791	0.772
		SynMOBIUS	0.794		0.808		0.811		0.674		0.836		0.885
5	UNet++_Binary	MOBIUS	0.754		0.851		0.720		0.643		0.832		0.878
		SMD+SLD	0.660	0.728	0.690	0.784	0.668	0.716	0.513	0.599	0.746	0.806	0.762
		SynMOBIUS	0.781		0.834		0.769		0.666		0.848		0.893
6	AEOS	MOBIUS	0.772		0.914		0.714		0.683		0.871		0.921
		SMD+SLD	0.612	0.720	0.731	0.847	0.589	0.677	0.511	0.622	0.731	0.822	0.763
		SynMOBIUS	0.809		0.927		0.749		0.712		0.883		0.941
7	SEG-U-Sclera	MOBIUS	0.642		0.655		0.669		0.512		0.716		0.642
		SMD+SLD	0.650	0.653	0.569	0.621	0.782	0.725	0.521	0.524	0.685	0.709	0.593
		SynMOBIUS	0.669		0.644		0.733		0.541		0.728		0.666
8	SAM-Iris	MOBIUS	0.633		0.639		0.657		0.487		0.675		0.683
		SMD+SLD	0.627	0.640	0.574	0.619	0.716	0.693	0.489	0.497	0.658	0.677	0.631
		SynMOBIUS	0.662		0.648		0.710		0.516		0.700		0.707
9	ShapeGAN-DLV3+	MOBIUS	0.592		0.743		0.539		0.482		0.729		0.779
		SMD+SLD	0.401	0.525	0.430	0.612	0.517	0.545	0.275	0.396	0.483	0.629	0.492
		SynMOBIUS	0.652		0.813		0.583		0.534		0.755		0.817

The probabilistic results for KU-CVML are not reported due to issues in the submission.

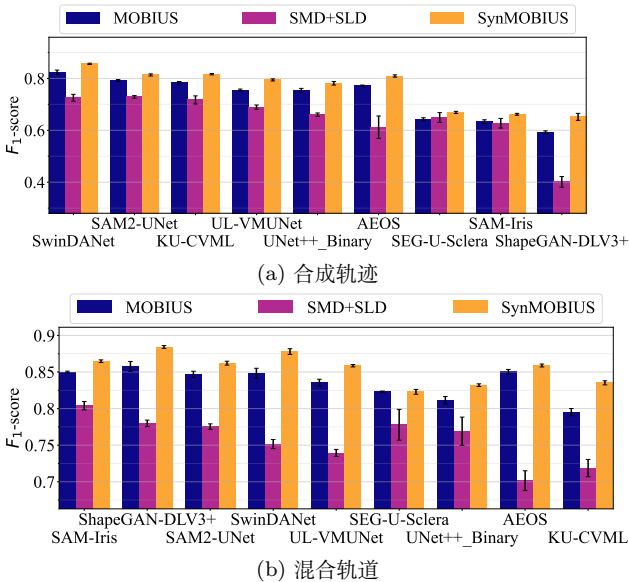


Figure 6. 对比在 (a) 合成数据和 (b) 混合训练数据上训练的提交作品，在评估数据集的所有图像上获得的 F_1 分数方面的性能表现。

合成数据赛道。由于 SSBC 2025 关注在模型开发中使用合成生成的数据，比赛的第一个赛道对仅使用

合成数据训练的模型进行基准测试。该赛道的结果在 Fig. 6a and Table 3 中展示。合成数据赛道的赢家是 SwinDANet，其在所有计算的性能指标上都优于其他方法。然而，大多数模型达到了相当有竞争力的结果，除 3 个提交之外，所有其他提交的得分都在 0.72–0.8 F_1 范围内。

Fig. 7 的第一行展示了模型在灰度预测中对二值化阈值的不同选择下性能如何变化。从 Table 3 可以观察到，几种模型在它们的二值化 F_1 分数与从灰度预测中获得的最佳 F_1 分数之间表现出显著的差异。这意味着模型可能会从采用专门的二值化阈值选择算法中大大受益。

有趣的是，模型的性能普遍在 MOBIUS 数据和类似 MOBIUS 的合成数据 (SynMOBIUS) 之间匹配良好，这表明生成的合成图像与原始实际数据之间存在相当大的对应关系。对于表现最好的模型，分数超过 F_1 0.8，表明即使与测试数据特征相当不同的合成数据中也可以学习到具有竞争力的分割模型。观察 SMD+SLD 的性能时，我们发现所有模型的结果与在 MOBIUS 或 SynMOBIUS 上的相应性能相比明显较弱，这表明合成训练数据与 SMD+SLD 图像之间的领域转移显著影响了分割结果——即使对于表现最好的模型也是如此。

混合赛项。在混合赛项中，参与者可以同时使用合成

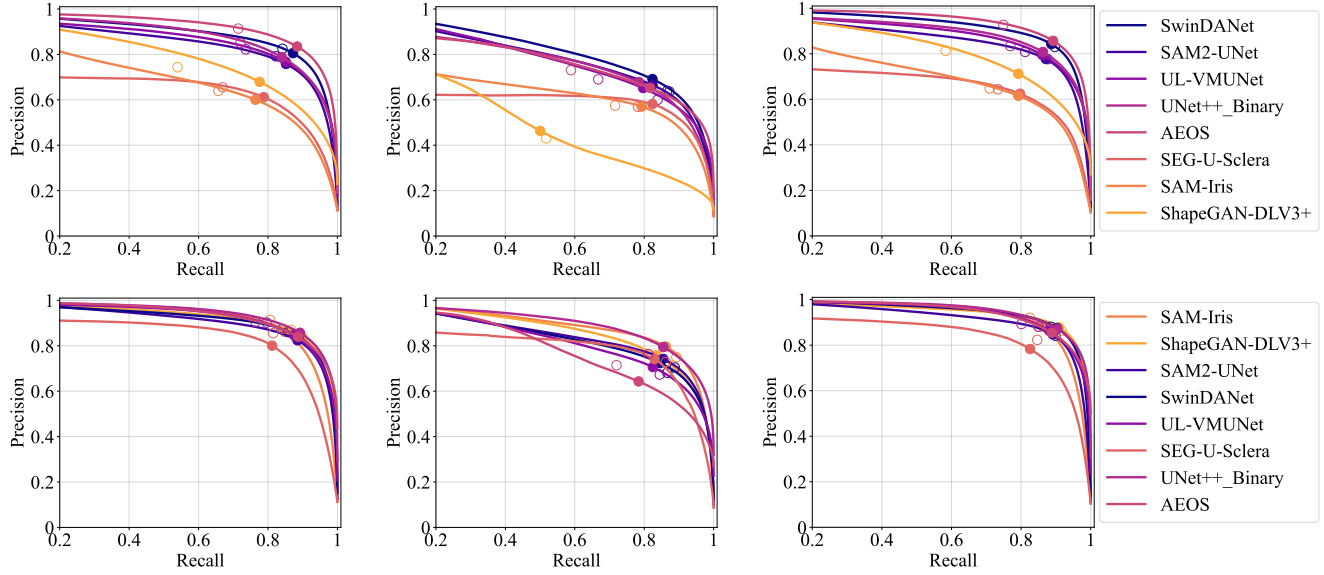


Figure 7. 提交的 SSBC 2025 模型的精确率-召回率曲线。用实心圆表示的操作点代表最佳可能的 F_1 得分 (F_1^{opt})，而空心圆表示由二元掩码产生的精确率-召回率点。第一行显示了在合成数据上训练的提交结果，而第二行包含了在混合数据场景中训练的提交结果。列中依次展示了在以下评估数据集上的结果：MOBIUS、SMD+SLD 和 SynMOBIUS。图例按相应赛道的排名排序。该图在彩色和放大查看时效果最佳。

Table 4. 对混合数据训练的模型进行比较评估。提交结果根据在三个评估数据集上达到的 F_1 分数的调和平均进行排名。 F_1 、精度、召回率和 IoU 分数是从提交的二进制掩码中计算得出的。精度-召回曲线上的最佳 F_1 分数 (F_1^{opt}) 和 AUC 值是从概率分割预测中计算得出的。在个别数据集结果旁边的列中也报告了其余性能指标的调和平均值。请注意，排名在性能指标之间相当一致。

Rank	Segmentation Model	Evaluation Dataset	From binary masks							From probabilistic predictions		
			F_1		Precision		Recall		IoU	F_1^{opt}	AUC	
1	SAM-Iris	MOBIUS	0.850		0.914		0.806		0.751	0.876	0.865	0.918
		SMD+SLD	0.804	0.839	0.749	0.853	0.893	0.840	0.684	0.833	0.881	0.910
		SynMOBIUS	0.865		0.920		0.825		0.769	0.889		0.933
1	ShapeGAN-DLV3+	MOBIUS	0.858		0.870		0.867		0.769	0.882		0.926
		SMD+SLD	0.780	0.838	0.712	0.808	0.885	0.889	0.649	0.811	0.863	0.909
		SynMOBIUS	0.884		0.862		0.917		0.800	0.901		0.944
3	SAM2-UNet	MOBIUS	0.847		0.860		0.858		0.756	0.871		0.906
		SMD+SLD	0.775	0.826	0.704	0.795	0.888	0.881	0.649	0.806	0.852	0.887
		SynMOBIUS	0.862		0.841		0.899		0.767	0.882		0.918
4	SwinDANet	MOBIUS	0.848		0.873		0.842		0.756	0.875		0.916
		SMD+SLD	0.752	0.822	0.681	0.800	0.866	0.864	0.623	0.786	0.850	0.896
		SynMOBIUS	0.878		0.881		0.886		0.790	0.896		0.945
5	UL-VMUNet	MOBIUS	0.835		0.902		0.796		0.732	0.873		0.927
		SMD+SLD	0.739	0.808	0.673	0.804	0.845	0.830	0.600	0.770	0.838	0.893
		SynMOBIUS	0.859		0.881		0.850		0.760	0.881		0.942
5	SEG-U-Sclera	MOBIUS	0.823		0.855		0.815		0.722	0.843		0.834
		SMD+SLD	0.778	0.807	0.729	0.799	0.862	0.841	0.659	0.811	0.833	0.820
		SynMOBIUS	0.823		0.823		0.846		0.721	0.846		0.840
7	UNet++_Binary	MOBIUS	0.811		0.906		0.764		0.710	0.889		0.941
		SMD+SLD	0.769	0.803	0.764	0.849	0.812	0.792	0.641	0.835	0.872	0.928
		SynMOBIUS	0.832		0.893		0.801		0.730	0.895		0.954
8	AEOS	MOBIUS	0.850		0.875		0.842		0.754	0.874		0.934
		SMD+SLD	0.701	0.797	0.714	0.816	0.720	0.799	0.569	0.733	0.823	0.880
		SynMOBIUS	0.859		0.881		0.850		0.762	0.880		0.945
9	KU-CVML	MOBIUS	0.795		0.725		0.928		0.689	-	-	-
		SMD+SLD	0.719	0.780	0.601	0.688	0.928	0.937	0.576	-	-	-
		SynMOBIUS	0.835		0.758		0.954		0.730	-	-	-

The probabilistic results for KU-CVML are not reported due to issues in the submission.

数据和真实世界数据进行模型训练。正如从 Fig. 6b and Table 4 中可以看到，这个赛项的结果（考虑最

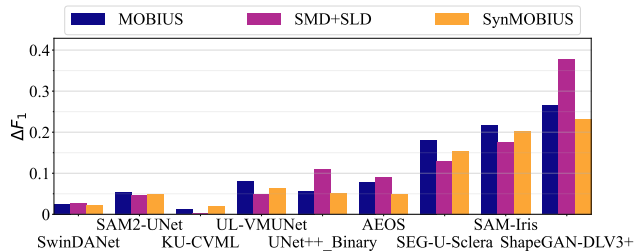


Figure 8. 提交模型从 Synthetic 轨道到 Mixed 轨道在三个测试数据集上的性能差异，即 MOBIUS、SMD+SLD 和 SynMOBIUS。图中展示了在达到的 F_1 评分上的差异，即 $F_1^{mixed} - F_1^{synthetic}$ 。正值表示在 Mixed 轨道上的更好性能。

强和最弱的模型)比合成赛项的更接近。这表明，与仅使用合成数据相比，使用真实和合成数据的混合来训练一个表现良好的模型仍然更加容易，因为需要在学习过程中加入专门的机制，来处理与真实世界数据相关的领域转移。混合赛项的获胜者 SAM-Iris 和亚军 ShapeGAN-DLV3+，在 F_1 分数上彼此之间相差 0.001，这在误差范围内。因此，他们被认为是 SSBC 2025 混合赛项的共同赢家。值得注意的是，所有方法都表现出色， F_1 分数在 0.78–0.84 的范围内。

有趣的是，来自合成赛道的两个表现最差的在混合赛道中表现最好，这主要是由于他们在 SMD+SLD 评估数据上显著的性能提升。来自合成赛道的两个表现最好的保持了他们的高性能，分别在混合赛道中获得了 3rd 和 4th 名。

Fig. 7 的第二行显示了灰度预测上的混合轨道结果。我们观察到，与第一行的合成轨道相比，精准度-召回率曲线明显更加紧密，这再次表明提交模型的性能更加接近。二进制 F_1 和最佳灰度 F_1 分数之间的差异仍然存在，但显著降低。这意味着对于混合训练的模型来说，一个好的阈值策略的重要性较低。例外是 UNet++_Binary，它在概率度量中获得了最高性能，但在二进制排名中仅获得了 7th 的名次。

6.2. 差异化性能

SSBC 2025 的目标是确定在开发巩膜分割模型时，使用合成数据的可行性。因此，我们感兴趣的是在合成数据上训练的模型与其混合训练的对应模型之间的性能差异。

从 Fig. 8 中可以看出，所有模型的性能从合成数据集到混合数据集都有所提升。然而，对于大多数模型而言，这种性能提升幅度较小，在 F_1 分数上的提升大多低于 0.05。然而，对于合成数据集表现最差的三个模型来说，加入真实世界训练数据后的性能提升幅度很大。这意味着在仅使用合成数据进行训练时，模型架构的选择非常重要，因为某些模型架构在没有真实数据来补充合成数据时表现会显著较差。然而，由于许多性能提升幅度较小，合成训练数据在隐私保护至关重要的情况下被证明是训练大多数模型的一种可行方法。

性能差异最大的是 ShapeGAN-DLV3+ 模型，该模

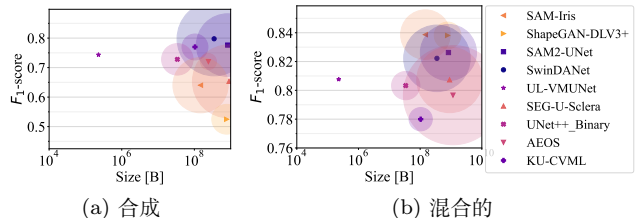


Figure 9. 性能/复杂性权衡。圆圈的面积对应于模型的计算复杂性，用浮点运算 (FLOPs) 表示。建议使用彩色查看并放大。

型具有生成器-判别器架构，尽管与经典 GAN 不同，因为判别器附加在生成器瓶颈处。由于众所周知 GAN 需要大量的训练数据，我们可以部分将性能差异归因于通过增加真实样本来增大的训练数据集。然而，我们注意到真实的 SBVPI 数据集比合成的训练数据集要明显小得多。第二大差异来自于 SAM-Iris 模型，其依赖于从合成训练数据到真实训练数据的显式过渡的训练流程，这解释了性能差异。同样，AEOS (导致了第四大差异) 在两个轨道上采用了完全不同的架构。因此，我们注意到合成轨道与混合轨道之间的四大性能差异中的三大差异是由于显式的技术选择而非领域变化造成的。

6.3. 不同复杂度下的表现

最后，我们研究了模型性能和复杂性之间的权衡，这对于现实世界的部署是一个关键因素。从 Fig. 9 可以看到，大多数提交的模型复杂性大致相当。然而，这里一个显著的例外是 UL-VMUNet，该模型在体积 (229 KB) 和计算复杂性 (60 MFLOPs) 相对较小的情况下实现了有竞争力的性能。这个结果与最近在巩膜生物识别方面的工作 [56,60] 一致，研究表明，通常使用的网络架构在不显著降低性能的情况下，可以大幅减小体积和复杂性。

还值得注意的是，UL-VMUNet 在合成数据上的排名高于在混合数据上的排名，这可能意味着，较小且不复杂的模型在使用合成数据进行训练时可以更好地泛化，并且不容易被数据生成过程产生的伪影干扰。

6.4. 定性评价

在 Fig. 10 中，我们展示了生成的分割掩码的一些最佳和最差情况。注意，即使在表现最好的样本中，模型常常在处理细微伪影时感到困难，例如由于虹膜上的镜面反射导致的伪影。在表现差的样本中，我们经常发现图像中的巩膜被遮挡，不论是被眼睑遮挡 (在部分闭合的眼睛情况下，例如在 Fig. 10 的第 4 行) 或被外部因素遮挡 (例如在 Fig. 10 的第 6 行的手指)。另一个错误来源是光照引起的皮肤颜色变化 (Fig. 10 的第 5 行)，尤其是当眼周的皮肤明显比其他地方的皮肤浅时。此外，光线不好的巩膜 (Fig. 10 的第 6 行) 也可能导致许多模型在分割时出现问题。

2025 年巩膜分割基准竞赛 (SSBC 2025) 被组织起来，用于评测现代分割模型在巩膜分割任务中的性能，

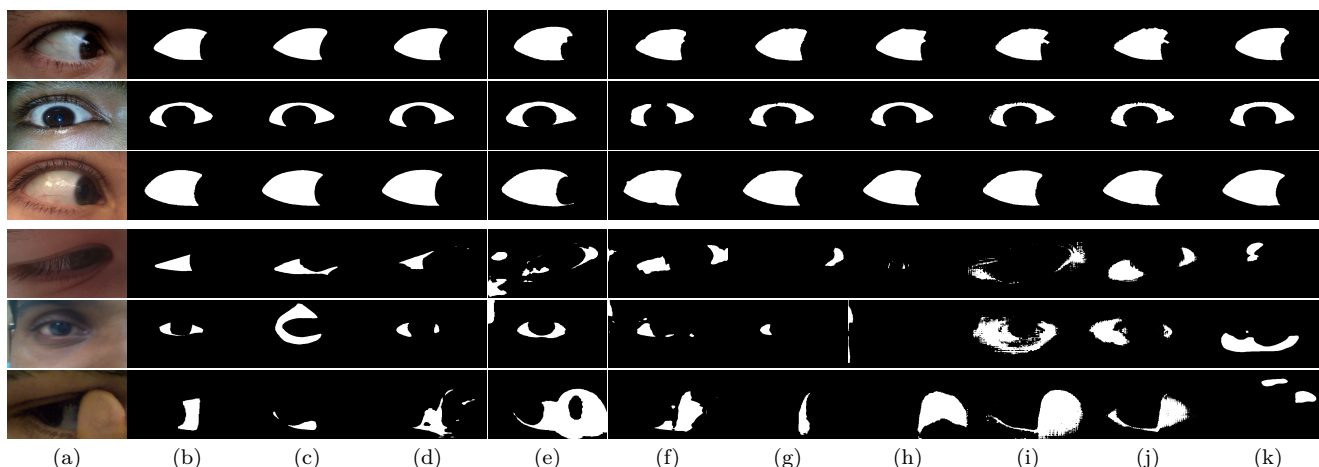


Figure 10. 对所选图像中的提交模型（就二进制掩码而言）进行定性比较。前三行包含来自评估数据集（依次为 MOBIUS、SMD+SLD、SynMOBIUS）的表现良好的样本，而后三行显示表现差的样本。观察经过评估的模型在分割质量上的差异。图中显示 (a) 原始图像；(b) 真实掩码；以及提交的二进制掩码来自：(c) SwinDANet, (d) SAM2-UNet, (e) KU-CVML, (f) UL-VMUNet, (g) UNet++_Binary, (h) AEOS, (i) SEG-U-Sclera, (j) SAM-Iris, 和 (k) ShapeGAN-DLV3+。

并确立使用合成生成数据开发和训练模型的可行性。使用合成数据确保训练分割模型不需要任何识别信息，并促进生物识别系统的伦理开发。为此，SSBC 2025 分两个赛道进行，其不同之处在于参赛者使用的训练数据——第一赛道使用纯合成数据，第二赛道使用合成数据和现实世界数据混合。九个研究组参加了此次竞赛。

合成赛道的获胜者是由 Couger Inc. 团队提交的 SwinDANet，而混合赛道则有两个共同获胜者，他们的结果相差不超过 0.001：SAM-Iris (Idiap – HES-SO 团队) 和 ShapeGAN-DLV3+ (AU 团队)。比赛结果表明合成数据具有高度真实度，并在模型训练中具有可行性，因为大多数方法在合成赛道和混合赛道中表现大致相同，而在两赛道之间显著提高性能模型，主要是由于方法上的决策，而不是合成训练数据中的任何不真实性或不当之处。这一结果对未来的研究来说非常重要，因为合成生成的数据集使得生物识别研究可以在没有隐私泄露或滥用风险的情况下进行，并且更易于获取或编译。

部分由 ARIS 研究计划 P2-0250 计量和生物识别系统、P2-0214 计算机视觉、ARIS 青年研究员计划以及 SNSF 项目 214653 FairMI 和 213369 StreamKG 支持。

References

- [1] S. Alkassar, W.-L. Woo, S. Dlay, and J. Chambers. Sclera recognition: on the quality measure and segmentation of degraded images captured under relaxed imaging conditions. *IET Biometrics*, 6(4):266–275, 2016. 1
- [2] S. Alkassar, W. L. Woo, S. S. Dlay, and J. A. Chambers. Robust sclera recognition system with novel sclera segmentation and validation techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(3):474–486, 2015. 1
- [3] S. Borse, Y. Wang, Y. Zhang, and F. Porikli. Inverse-form: A loss function for structured boundary-aware segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5901–5911, 2021. 5
- [4] K. Boyd, K. H. Eng, and C. D. Page. Area under the precision-recall curve: point estimates and confidence intervals. In *Springer Joint European conference on machine learning and knowledge discovery in databases (ECML PKDD)*, pages 451–466, 2013. 5
- [5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. DeepLab: Semantic Image Segmentation With Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(4):834–848, 2018. 5
- [6] A. Das. Towards Multi-modal Sclera and Iris Biometric Recognition with Adaptive Liveness Detection. PhD thesis, Griffith University, 2017. 2
- [7] A. Das, S. Atreya, A. Mukherjee, M. Vitek, H. Li, C. Wang, G. Zhao, F. Boutros, P. Siebke, J. N. Kolf, et al. Sclera segmentation and joint recognition benchmarking competition: Ssrbc 2023. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2023. 2
- [8] A. Das, R. Kunwar, U. Pal, M. A. Ferrer, and M. Blumenstein. An online learning-based adaptive biometric system. In *Adaptive Biometric Systems*, pages 73–96. Springer, 2015. 1
- [9] A. Das, P. Mondal, U. Pal, M. Blumenstein, and M. A. Ferrer. Sclera vessel pattern synthesis based on a non-parametric texture synthesis technique. In *Springer International Conference on Computer Vision and Image Processing (CVIP)*, pages 241–250, 2017. 1
- [10] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein. Sclera recognition using dense-sift. In *IEEE International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 74–79, 2013. 1

- [11] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein. A new efficient and adaptive sclera recognition system. In *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 1–8, 2014. 1
- [12] A. Das, U. Pal, M. Blumenstein, and M. A. F. Ballester. Sclera recognition-a survey. In *IEEE IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 917–921, 2013. 1
- [13] A. Das, U. Pal, M. Blumenstein, C. Wang, Y. He, Y. Zhu, and Z. Sun. Sclera segmentation benchmarking competition in cross-resolution environment. In *IEEE/IAPR International Conference on Biometrics*, 2019. 2
- [14] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. SSBC 2015: Sclera Segmentation Benchmarking Competition. In *Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 742–747, 2015. 2
- [15] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. A framework for liveness detection for direct attacks in the visible spectrum for multimodal ocular biometrics. *Elsevier Pattern Recognition Letters*, 82:232–241, 2016. 1
- [16] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. SS-RBC 2016: sclera segmentation and recognition benchmarking competition. In *International Conference on Biometrics (ICB)*, pages 1–6, 2016. 2
- [17] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. A decision-level fusion strategy for multimodal ocular biometric in visible spectrum based on posterior probability. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 794–798, 2017. 1
- [18] A. Das, U. Pal, M. A. Ferrer, M. Blumenstein, D. Štepec, P. Rot, Z. Emeršič, P. Peer, V. Štruc, and S. Kumar. SSERBC 2017: Sclera segmentation and eye recognition benchmarking competition. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 742–747, 2017. 2
- [19] A. Das, U. Pal, M. A. Ferrer, M. Blumenstein, D. Štepec, P. Rot, P. Peer, and V. Štruc. SSBC 2018: Sclera Segmentation Benchmarking Competition. In *International Conference on Biometrics (ICB)*, pages 303–308, 2018. 2
- [20] S. Das, I. De Ghosh, and A. Chattopadhyay. An efficient deep sclera recognition framework with novel sclera segmentation, vessel extraction and gaze detection. *Signal Processing: Image Communication*, 97:116349, 2021. 1
- [21] S. Das, I. De Ghosh, and A. Chattopadhyay. Sclera biometrics in restricted and unrestricted environment with cross dataset evaluation. *Displays*, 74:102257, 2022. 1
- [22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009. 5
- [23] Ž. Emeršič, L. L. Gabriel, V. Štruc, and P. Peer. Pixel-wise ear detection with convolutional encoder-decoder networks. *IET Biometrics*, 2017. 5
- [24] M. B. Erden, S. Cansiz, O. Caki, H. Khattak, D. Etiz, M. C. Yakar, K. Duruer, B. Barut, and C. Gunduz-Demir. Fourierloss: Shape-aware loss function with fourier descriptors. *Neurocomputing*, 638:130155, 2025. 5
- [25] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi. Openeds: Open eye dataset. *arXiv preprint arXiv:1905.03702*, 2019. 1
- [26] V. Gottemukkula, S. Saripalle, S. P. Tankasala, and R. Derakhshani. Method for using visible ocular vasculature for mobile biometrics. *IET Biometrics*, 5(1):3–12, 2016. 1
- [27] A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 5
- [28] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 4
- [29] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 5
- [30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017. 5
- [31] P. Kansal and S. Devanathan. Eyenet: Attention based convolutional encoder-decoder network for eye region segmentation. In *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3688–3693, 2019. 5
- [32] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:12104–12114, 2020. 3
- [33] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8110–8119, 2020. 3, 4
- [34] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3, 4
- [35] A. Kirillov, E. Mintun, N. Ravi, H. Mao, L. Rolland, R. Gustafson, C. Xiao, S. Whitehead, A. Cevallos, Z. Howald, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. 5
- [36] R. Liu, J. Lehman, et al. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. 5
- [37] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, et al. Swin transformer

- v2: Scaling up capacity and resolution. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12009–12019, 2022. 5
- [38] J. Lozej, B. Meden, V. Štruc, and P. Peer. End-to-end iris segmentation using U-Net. In IEEE International Work Conference on Bioinspired Intelligence (IWOBI), pages 1–6, 2018. 5
- [39] J. Lozej, D. Štepec, V. Štruc, and P. Peer. Influence of segmentation on deep iris recognition performance. In 2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), pages 1–6, 2019. 5
- [40] D. M. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2:37–63, 2011. 5
- [41] P. Radu, J. Ferryman, and P. Wild. A robust sclera segmentation algorithm. In IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS), pages 1–6, 2015. 1
- [42] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, et al. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024. 5
- [43] D. Riccio, N. Brancati, M. Frucci, and D. Gagnaniello. An unsupervised approach for eye sclera segmentation. In Springer Iberoamerican Congress on Pattern Recognition, pages 550–557, 2017. 1
- [44] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In Springer International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 234–241, 2015. 5
- [45] P. Rot, Ž. Emeršič, V. Štruc, and P. Peer. Deep multi-class eye segmentation for ocular biometrics. In IEEE International Work Conference on Bioinspired Intelligence (IWOBI), pages 1–8, 2018. 5
- [46] P. Rot, M. Vitek, K. Grm, Ž. Emeršič, P. Peer, and V. Štruc. Deep sclera segmentation and recognition. In *Handbook of vascular biometrics*, pages 395–432. Springer, 2020. 1
- [47] C. Ryali, Y.-T. Hu, D. Bolya, C. Wei, H. Fan, P.-Y. Huang, V. Aggarwal, A. Chowdhury, O. Poursaeed, J. Hoffman, et al. Hiera: A hierarchical vision transformer without the bells-and-whistles. In PMLR International Conference on Machine Learning (ICML), pages 29441–29454, 2023. 5
- [48] T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015. 5
- [49] A. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigun, K. Raja, R. Raghavendra, C. Busch, T. F. Pereira, et al. Cross-eyed 2017: Cross-spectral iris/periocular recognition database and competition. In IEEE International Joint Conference on Biometrics (IJCB), 2017. 3
- [50] M. Tan and Q. V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In International Conference on Machine Learning (ICML), 2019. 5
- [51] D. Tomašević, F. Boutros, N. Damer, P. Peer, and V. Štruc. Generating bimodal privacy-preserving data for face recognition. *Engineering Applications of Artificial Intelligence (EAAI)*, 133:108495, 2024. 1
- [52] D. Tomašević, F. Boutros, C. Lin, N. Damer, V. Štruc, and P. Peer. ID-Booth: Identity-consistent face generation with diffusion models. arXiv preprint arXiv:2504.07392, 2025. 1
- [53] D. Tomašević, P. Peer, and V. Štruc. BiOcularGAN: Bimodal synthesis and annotation of ocular images. In IEEE International Joint Conference on Biometrics (IJCB), pages 1–10, 2022. 1, 3, 4
- [54] D. Tomašević, P. Peer, and V. Štruc. BiFaceGAN: Bimodal Face Image Synthesis. In *Face Recognition Across the Imaging Spectrum (FRAIS)*, pages 273–311. Springer, 2024. 1
- [55] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Bochoon, and S. Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 969–977, 2018. 5
- [56] M. Vitek, M. Bizjak, P. Peer, and V. Štruc. Ipad: Iterative pruning with activation deviation for sclera biometrics. *Journal of King Saud University-Computer and Information Sciences*, 35(8):101630, 2023. 1, 8
- [57] M. Vitek, A. Das, D. R. Lucio, L. A. Zanolensi, D. Menotti, J. N. Khiarak, M. A. Shahpar, M. Asgari-Chenaghlu, F. Jaryani, J. E. Tapia, et al. Exploring bias in sclera segmentation models: A group evaluation approach. *IEEE Transactions on Information Forensics and Security (TIFS)*, 18:190–205, 2022. 1, 2
- [58] M. Vitek, A. Das, Y. Pourcenoix, A. Missler, C. Paumier, S. Das, I. De Ghosh, D. R. Lucio, L. A. Zanolensi Jr., D. Menotti, F. Boutros, N. Damer, J. H. Grebe, A. Kuijper, J. Hu, Y. He, C. Wang, H. Liu, Y. Wang, Z. Sun, D. Osorio-Roig, C. Rathgeb, C. Busch, J. Tapia Farias, A. Valenzuela, G. Zampoukis, L. Tsochatzidis, I. Pratikakis, S. Nathan, R. Suganya, V. Mehta, A. Dhall, K. Raja, G. Gupta, J. N. Khiarak, M. Akbari-Shahper, F. Jaryani, M. Asgari-Chenaghlu, R. Vyas, S. Dakshit, S. Dakshit, P. Peer, U. Pal, and V. Štruc. SSBC 2020: Sclera segmentation benchmarking competition in the mobile environment. In IEEE International Joint Conference on Biometrics (IJCB), 2020. 2, 3, 4
- [59] M. Vitek, P. Rot, V. Štruc, and P. Peer. A comprehensive investigation into sclera biometrics: a novel dataset and performance study. *Springer Neural Computing and Applications*, pages 1–15, 2020. 1, 2
- [60] M. Vitek, V. Štruc, and P. Peer. Gazenet: A lightweight multitask sclera feature extractor. *Alexandria Engineering Journal*, 112:661–671, 2025. 1, 8

- [61] R. Wu, Y. Liu, P. Liang, and Q. Chang. Ultralight vm-unet: Parallel vision mamba significantly reduces parameters for skin lesion segmentation. arXiv preprint arXiv:2403.20035, 2024. 5
- [62] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021. 5
- [63] X. Xiong, Z. Wu, S. Tan, W. Li, F. Tang, Y. Chen, S. Li, J. Ma, and G. Li. Sam2-unet: Segment anything 2 makes strong encoder for natural and medical image segmentation. ArXiv, abs/2408.08870, 2024. 5
- [64] Y. Zhang, H. Ling, J. Gao, K. Yin, J.-F. Lafleche, A. Barriuso, A. Torralba, and S. Fidler. Datasetgan: Efficient labeled data factory with minimal human effort. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10145–10155, 2021. 3
- [65] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp. A new human identification method: Sclera recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(3):571–583, 2011. 1
- [66] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, 2018. 5
- [67] X. Zhu, H. Hu, S. Lin, and J. Dai. Deformable convolutional networks v2: More deformable, better results. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5