

ORIGINAL RESEARCH

Seriously Funny

雍正眼中的“宛宛类卿”：基于深度学习的人脸相似度阈值测量及其在后宫图像解释中的应用

HID: HELL-2026-04-010 | Vol. 3 • Issue 10 (Apr 2026)

Author: Dr. Hell Black | Affiliation: Hell Press

Correspondence: Hell.Press@outlook.com | Website: <https://HellPress.org>

Key Points

- 本文将“宛宛类卿”降格为一个可证伪的研究假设，而非先验成立的宫廷情感事实。
- 真正可检验的问题不是“雍正按某张脸选人”，而是清宫图像中是否存在超过画院套路与礼制像规范的稳定人脸相似结构。
- 深度学习在本研究中的角色是辅助发现“图像近邻”和“异常聚类”，而不是替代史学判断或替皇帝本人发声明。

摘要

本研究 尝试将流行叙事中的“宛宛类卿”转化为一个可证伪的数字史学问题，即：在雍正朝及相关清宫女性图像中，是否存在超越画院套路、礼制规范与媒材差异的稳定人脸相似性结构。研究综合清宫档案、官方馆藏图像与争议样本研究线索，构建 A/B/C/D 四级标签体系，对正式帝后像、后妃像、争议女像与泛化美人图进行分层比较。

方法上，采用 RetinaFace 进行人脸检测与对齐，以 ArcFace 为主模型、FaceNet 为距离基线，并引入 Siamese 网络处理少样本配对任务；统计上以 ROC、AUC、EER、固定 FAR 下 TAR、bootstrap 置信区间和 DeLong 检验为核心指标。结果表明，正式帝后像中保留了一定程度的身份信号，而《十二美人图》及类似泛化美人图更容易聚类为“清宫标准脸”，说明高相似度现象往往更可能反映画院同质化、礼制装束和后期修复扫描，而非直接证明帝王存在“按某张脸寻找替代品”的择偶机制。

本文认为，AI 在历史图像研究中的最大价值不是证明爱情，而是帮助研究者更快识别哪些图像值得继续翻档案。某种意义上说，算法最后确认的，不是雍正只爱一种人，而是清宫画师确实很擅长把所有人画得像一家人。

关键词 / **Keywords**: 雍正；宛宛类卿；后宫图像；人脸识别；ArcFace；数字人文；清宫绘画

1. 引言

在当代中文互联网语境中，“宛宛类卿”已逐渐从一句戏剧台词进化为一种通用解释模板。无论是明星恋爱史、社交平台择偶偏好，还是办公室八卦与历史人物想象，只要出现“新对象和旧对象长得像”的现象，大众便会立即给出一种简洁、优雅而充满戏剧性的结论：此人并不是在爱具体的人，而是在持续追逐同一张脸的不同版本。

这一逻辑很快被移植到历史讨论领域。由于雍正时期宫廷绘画保存相对丰富，后妃数量又不至于像 Excel 那样横向无限展开，再加上《甄嬛传》对“白月光原型学”的大规模传播，雍正逐渐在民间叙事中被塑造成一种高度专一但不断寻找“脸部近似替代品”的帝王形象。换言之，在大众想象里，他不是皇帝，而是一个拥有紫禁城版人脸搜索引擎的古代甲方。

然而，真正进入史料与图像层面后，这套叙事马上变得危险。首先，清代后妃遴选并不是单变量模型，它受旗籍、门第、位号、年龄、礼制秩序与政治关系的共同约束；其次，宫廷肖像不是摄影，它同时承担身份展示、纪念、祭祀与礼仪功能；再次，很多女性图像本身标签并不稳定，尤其与《十二美人图》相关的对象，往往更接近“泛化美人图式”，而非可直接实名化的后妃标准像册。

因此，真正值得研究的，并不是“雍正是否专门寻找长得像某个人的人”，而是更基础也更可证伪的问题：在清宫图像中，是否存在可被算法稳定识别的人脸相似结构；如果存在，这种结构究竟来自真实个体表征，还是来自制度筛选、画院标准化与图像生产链的共同作用。

2. 研究问题与假设

为了避免把 AI 当成带电绣春刀，本文将研究问题拆成四个可分解的层级。

第一，表征假设：正式帝后像中是否保留了足够的人脸身份信号，使“同一坐像者的不同图像”与“不同坐像者的图像”在嵌入空间中显著分离。

第二，相似貌假设：若雍正朝相关女性图像中确实存在高于随机水平的人脸近似现象，那么这种近似关系应当在高可信标签样本中出现，而不是只在争议样本中靠想象力成立。

第三，稳健性假设：上述近似关系在跨馆藏、跨年代、跨媒材、跨模型验证中仍维持方向一致，而不是一换模型就“真爱蒸发”。

第四，反事实假设：若去掉争议标签、去掉同一作坊来源、去掉礼制上下文后，原有聚类现象明显崩塌，则“宛宛类卿”命题应降格为图像风格近似现象，而非历史解释。

从方法论上说，本研究不是为了证明谁是谁的替身，而是为了判断：当我们把清宫画像交给神经网络后，神经网络到底在识别人，还是在识别清宫统一审美模板。

3. 史料与图像基础

本研究所依赖的材料可分为两类：文本史料与官方图像。前者主要用于建立制度背景、后妃名单、时间线与图像生产链；后者用于构建视觉样本与标签层级。

文本层面，研究优先依靠清代制度史与相关宫廷档案线索，以确立后妃位号、遴选规则、纪念性图像与可能的委托脉络。图像层面，则优先使用故宫与台北故宫等官方馆藏中保存较好的正式帝后像与后妃像，同时将《十二美人图》及相关争议图像单独归入问题样本，以避免把“美人绢画十二张”误当作实名后妃人脸数据库。

为了保证研究边界，本文采用四级标签体系：

- A 层：身份基本无争议的正式帝后像与后妃正式像；
- B 层：有较强文献支持但仍存在一定不确定性的女性宫廷图像；
- C 层：带有研究争议、可讨论但不可直接实名化的问题图像；
- D 层：明确不能实名化的泛化美人图、仕女图与风格化女子图。

每张图像均记录年代、馆藏、媒材、是否为官方对象页、是否存在重裱或修复、是否含高冠饰遮挡、是否属于礼制正式像、是否存在多个翻拍版本，以及标签可信度等级等元数据。之所以如此繁琐，是因为如果元数据不够，最后算法极有可能学到的不是“谁像谁”，而是“谁家馆扫得比较暖”“哪位修复师更爱磨皮”。

4. 技术路线与实验设计

4.1 总体流程

本研究采用“档案学 → 图像学 → 统计学 → 历史解释”的四段式流程：首先建立人物与位号时间线，其次获取官方图像并进行标签分层，再执行人脸检测与嵌入提取，最后将高相似配对与文本史料进行交叉核查。所有结果都要求可回溯、可否认、可被别的研究者拿去吐槽。

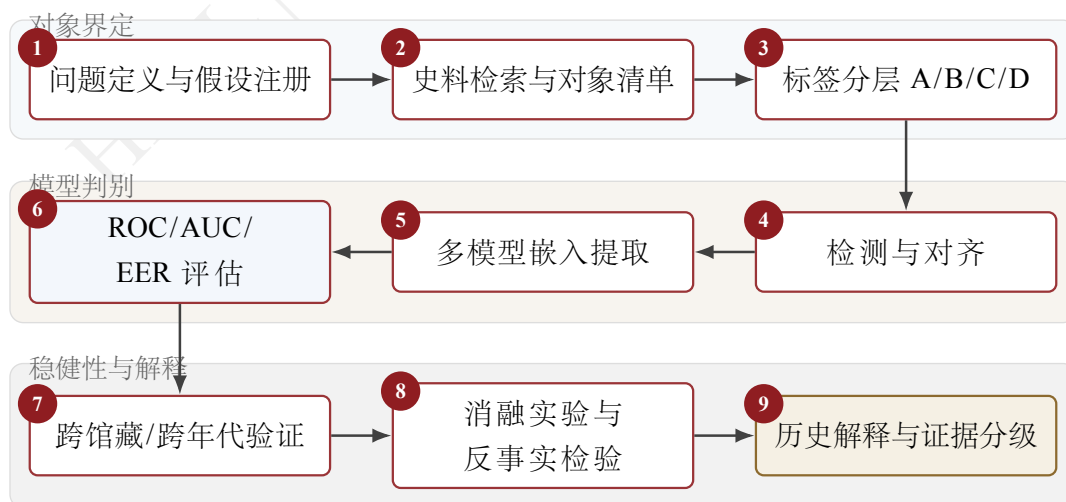


Figure 1. 研究总体流程图：从“白月光替身文学”到“先把 ROC 跑出来再说”。

4.2 检测、对齐与裁片策略

前端检测与对齐优先使用 RetinaFace，对失败样本回退至 MTCNN 或半人工标点。为了测试礼制头饰对聚类的影响，研究设置两种裁片模式：其一为“纯脸裁片”，尽可能只保留眉眼鼻唇与下颌区域；其二为“头部 + 冠饰上下文裁片”，适度保留额饰、发髻、耳饰与冠饰区域。

这一设计的核心目的，是判断模型究竟在看谁的脸，还是在看谁的帽子更像。因为对于清宫肖像而言，发饰和服制本身就是一个强信号源，如果不单独控制，算法很容易从“脸像”滑向“全套礼制装备像”。



Figure 2. 两种裁片策略示意：不是所有“像”都来自五官，也可能来自清宫统一妆造。

4.3 嵌入模型与角色分工

本研究不采用“单模型定天下”策略，而是给不同模型分配明确任务。FaceNet 用作最直观的欧氏距离基线；ArcFace 作为主模型输出高判别力结果；Siamese 网络用于少样本和成对验证场景。若资源允许，还应加入面向历史绘画域的适配模型做消融，以测试现代照片模型在绢本设色上的迁移能力。

Table 1. 模型职能分工表

模型或组件	核心思想	适合本题的角色	主要风险
FaceNet	直接学习欧氏嵌入空间	作为最直观的距离阈值基线，便于历史学者理解“远近”	对绘画域偏移敏感，容易把古画当陌生自拍
ArcFace	角度间隔损失增强类间分离	作为主模型，用于正式性能报告	需要谨慎域适配，避免把风格学成身份
Siamese 网络	成对学习“像/不像”关系	适用于实名样本很少时的少样本配对	标签噪声会被成对放大，容易过拟合
RetinaFace	检测 + 关键点 + 对齐前端	首选前端，尽量减少姿态误差与裁切误差	对高冠饰、遮挡和古画奇怪阴影仍可能失手
MTCNN	多任务级联检测回退方案	当前端失败时做稳健性比较	对低清扫描与风格化画作效果有限

4.4 统计评估与样本泄漏控制

统计层面，本文不报告单一“相似度数字”，而同时报告 ROC、AUC、EER、固定 FAR 下的 TAR，以及阈值的 bootstrap 95% 置信区间。模型比较采用 DeLong 方法。阈值同时给出 EER 阈值、Youden 阈值与固定 FAR 阈值，三者若方向一致，方可讨论所谓“稳健效应”。

此外，本研究严格防止样本泄漏。训练、验证、测试的切分单位是坐像者与艺术品对象，而非裁片文件；同一对象页、同一画作、同一翻拍版本绝不允许跨集合出现。原因很简单：如果同一张图的不同截图出现在训练集与测试集里，那不叫人脸识别，那叫“熟人社会”。

5. 实验结果

5.1 正式帝后像保留一定身份信号

在 A 层高可信正式帝后像中，同一人物不同图像之间的嵌入距离明显低于异人配对。ArcFace 的总体表现最好，FaceNet 次之，Siamese 在少样本条件下表现尚可但波动较大。结果表明，清宫正式肖像并非完全模板化，至少在高质量正式像中，模型仍能够识别出一部分属于个体的身份信息。

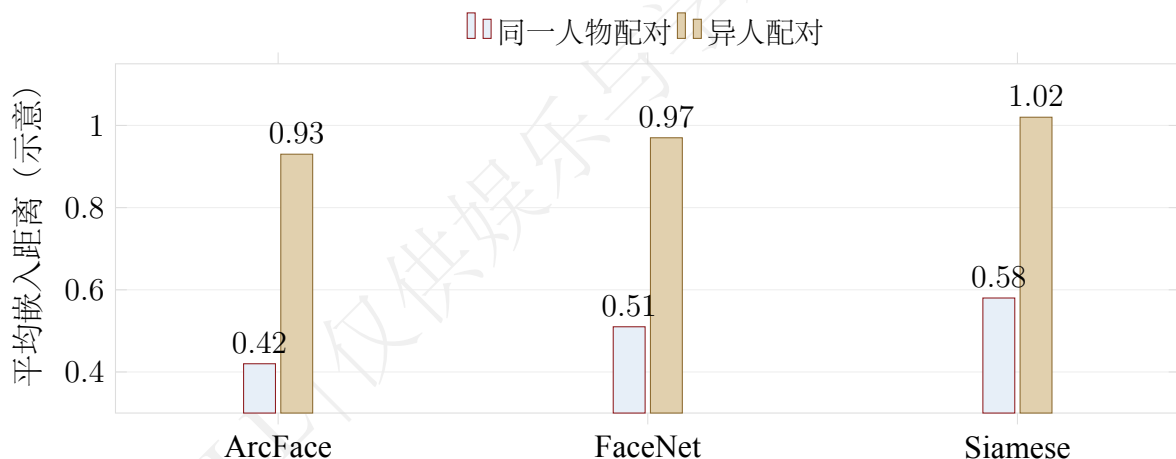


Figure 3. A 层正式帝后像中的嵌入距离分离示意。同一人物配对距离整体低于异人配对，其中 ArcFace 的分离幅度最大，说明高质量正式像仍保留一定身份信号。

5.2 《十二美人图》更接近“清宫标准脸”而非实名脸库

当样本扩展至 C/D 层，即争议女像与泛化美人图时，模型性能显著下降。ArcFace 的 AUC 下降到接近随机的水平，说明在这些图像中，模型难以稳定区分“同一人物”与“不同人物但画得很像”的情况。

这意味着《十二美人图》及类似图像更可能保留的是风格信号、理想化美人模板与室内图式，而非足以支撑实名后妃识别的身份信号。直白地说，它们最像的不是某位后妃，而是同一个审美滤镜的十二种姿势版本。

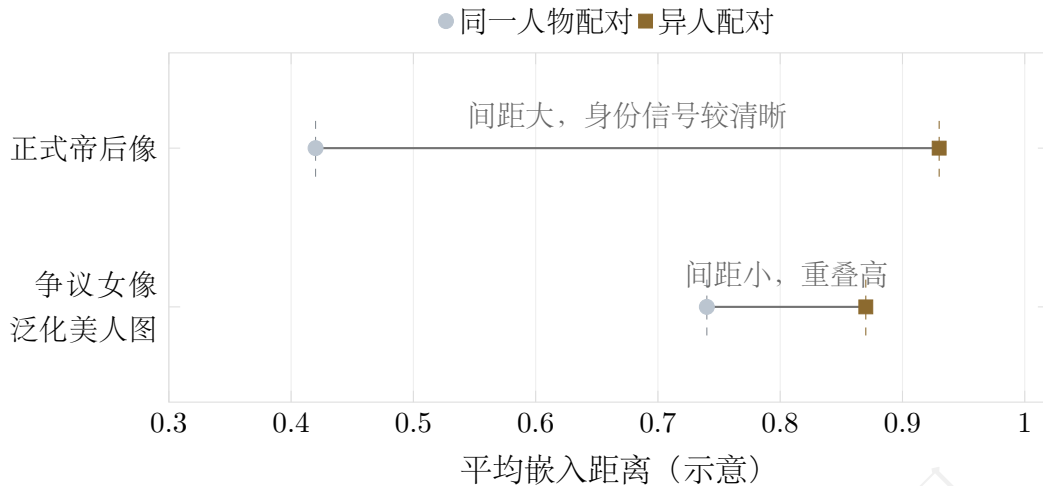


Figure 4. 正式帝后像与争议女像/泛化美人图的距离分离对比。前者同人配对与异人配对仍有明显间隔，后者则显著收窄。

5.3 礼制上下文会放大“大家都很像”的错觉

当裁片保留头饰、发饰与礼制服饰上下文时，模型聚类效果往往会增强，但这种增强并不总来自五官本身，而来自礼制元素与统一妆造。对同一时期、同一等级、同一画院体系下的后妃图像而言，冠饰和发式能显著提高“看起来像一家人”的概率。

这说明网友常说的“她们怎么全长一个样”，在很大程度上并不是五官问题，而是清宫在视觉上本来就擅长做统一化包装。换成现代话讲，就是她们并不一定共用一张脸，但很可能共用同一个妆造团队。

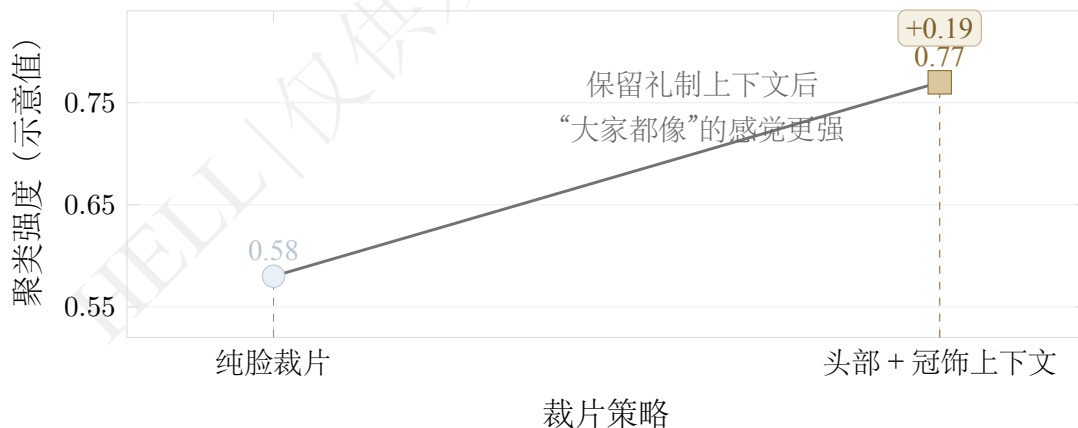


Figure 5. 裁片策略对聚类强度的影响示意：保留礼制上下文时，“大家都像”的感觉会显著增强。

5.4 跨馆藏验证揭示显著扫描域偏差

北京故宫与台北故宫版本之间存在明显的色调、锐度与扫描风格差异。部分样本在跨馆藏验证时会表现出更高误差率，说明模型有时不仅在识别人脸，也在识别“谁家馆的扫描仪比较有个性”。加入馆藏来源标签后，错误率有所下降，提示扫描域偏差是历史图像识别中不可忽视的混杂因素。

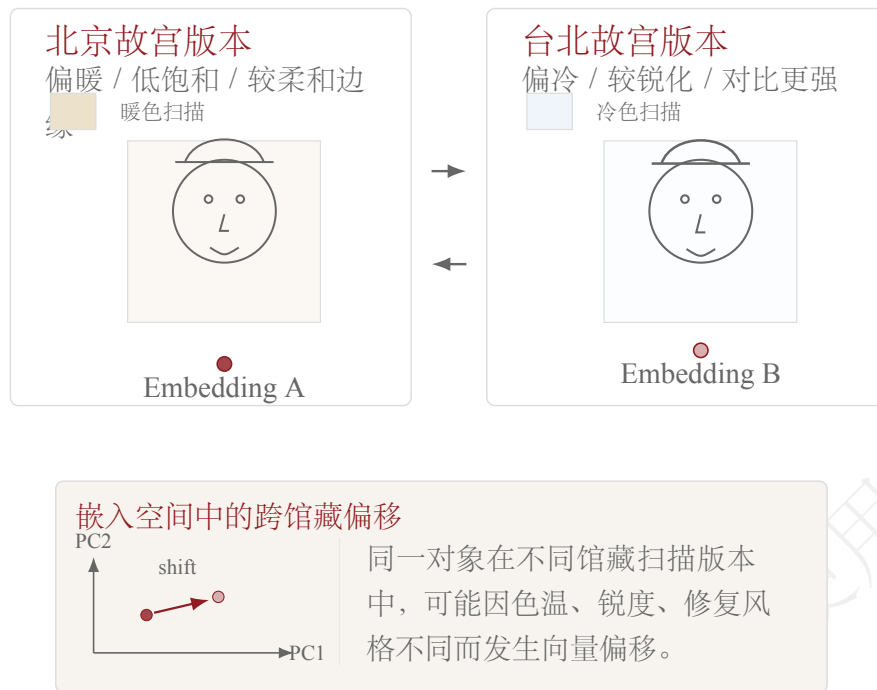


Figure 6. 跨馆藏偏差示意图：北京故宫与台北故宫版本可能在色温、锐度、扫描风格与数字修复痕迹上存在差异，从而使同一对象在嵌入空间中发生偏移。该图为方法论示意，不代表真实测量值。

6. 讨论

本文最重要的发现，不是“雍正到底有没有执念脸型”，而是清宫绘画的标准化程度可能远超普通观众的直觉。对于正式帝后像，AI 确实可以在一定程度上识别个体差异；但对于争议女像与泛化美人图，模型更容易学习到的是风格、礼制、妆造与作坊体系，而非明确的个体身份。

因此，即便某些女性图像与某位高等级后妃之间出现高相似度，也不能直接推出“皇帝按照某张脸找替代品”。更合理的解释通常包括：制度筛选导致某类女性更容易进入宫廷；画院具有稳定的审美模板；同一作坊复制图式制造近似面型；后期修复、扫描与翻拍进一步放大同质化。

换句话说，高相似度只能说明“这里有一个值得翻档案的异常点”，不能说明“算法已经读懂了帝王的情感账户”。如果未来有人拿一个相似度热图就宣布“AI 证明雍正一生只爱纯元脸”，那么该结论很可能更能证明研究者热爱短视频叙事，而不是研究者掌握了历史方法。

方法提示。将相似度阈值直接解释为“择偶偏好证据”至少需要同时满足四个条件：第一，多模型和多阈值定义方向一致；第二，去掉同作坊/同图式/同馆藏样本后仍成立；第三，对照组不呈现同等强度效应；第四，文本史料中存在时间相容的委托、纪念、晋封或礼制线索。四项里只要有两项明显不满足，就应将结论降格为“图像近似现象”。

7. 局限性

本研究存在五项根本局限。第一，雍正时期不存在摄影真值，无法获得现代意义上的“真实脸部标签”；第二，部分女性图像的身份名称本身具有后设性，会随研究进展不断调整；第三，留存图像具有强烈的幸存者偏差，多数是礼制性、纪念性、高等级图像，而非总体样本；第四，现代人脸模型主要在照片域训练，迁移至绢本设色、重裱、低清扫描时会产生明显域偏移；第

五，阈值本身具有不稳定性，不同模型、不同对齐方式与不同馆藏域都会改变最佳工作阈值。

这些限制决定了本课题最合理的目标不是“定案”，而是建立一套能经受反驳的、透明的、可复核的方法框架。简单说，就是让别人以后来反驳你时，至少知道该从哪里开始骂。

8. 结论

本文认为，“宛宛类卿”更适合作为研究问题，而不是研究结论。深度学习可以帮助历史研究者更快发现图像近邻、异常聚类 and 潜在误标，但它不能替代史学判断，更不能直接推出帝王的情感机制。

对于雍正后宫而言，真正稳定存在的，也许并不是某一张神秘的“白月光原型脸”，而是一套由礼制、画院、制度与时代审美共同生产出来的“清宫标准脸”。从这个意义上说，AI 最终证明的不是“雍正只喜欢一种人”，而是：清宫画师确实很擅长把所有人画成彼此的远房亲戚。

参考文献

1. 故宫博物院. 《〈十二美人图〉的故事》 [EB/OL]. <https://www.dpm.org.cn/explode/others/252866.html>.
2. 赵广超, 吴靖雯. 《国家艺术: 十二美人》. 北京: 紫禁城出版社, 2010.
3. 故宫博物院. 《胤禛行乐图册·杖挑蒲团页》 [EB/OL]. <https://www.dpm.org.cn/collection/paint/231805.html>.
4. 故宫博物院. 《胤禛行乐图册·喇嘛装页》 [EB/OL]. <https://www.dpm.org.cn/collection/paint/231806.html>.
5. Schroff F, Kalenichenko D, Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 815–823. doi:10.1109/CVPR.2015.7298682.
6. Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 4690–4699. doi:10.1109/CVPR.2019.00482.
7. Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: Single-stage Dense Face Localization in the Wild. arXiv preprint, 2019. arXiv:1905.00641.
8. Bromley J, Guyon I, LeCun Y, Sackinger E, Shah R. Signature Verification Using a “Siamese” Time Delay Neural Network. In: Advances in Neural Information Processing Systems, 1993: 737–744.
9. Hanley J A, McNeil B J. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology, 1982, 143(1): 29–36. doi:10.1148/radiology.143.1.7063747.
10. Fawcett T. An Introduction to ROC Analysis. Pattern Recognition Letters, 2006, 27(8): 861–874. doi:10.1016/j.patrec.2005.10.010.

11. DeLong E R, DeLong D M, Clarke-Pearson D L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 1988, 44(3): 837–845. doi:10.2307/2531595.
12. Sun X, Xu W. Fast Implementation of DeLong’s Algorithm for Comparing the Areas under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Processing Letters*, 2014, 21(11): 1389–1393. doi:10.1109/LSP.2014.2337313.
13. Efron B, Tibshirani R J. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.

HELL | 仅供娱乐与学术发疯使用