

Research on Multi-dimensional Quality Quantification Standards for Portrait Dataset Images Used in Face Recognition Algorithm Training

Yitong Fu^{1*}

¹ Macquarie University, Sydney, 2109, Australia

*jesg278@gmail.com

<https://doi.org/10.70695/IAAI202603A2>

Abstract

The training effectiveness of face recognition algorithms is related not only to model architecture and loss functions, but also to the image quality, annotation quality, distributional structure, and compliance governance of portrait datasets. To address the current problems of fragmented quality indicators, a high proportion of subjective screening, and the lack of unified standards for handling low-quality samples in portrait data collection and training admission, this paper proposes a multi-dimensional quality quantification standard for portrait images oriented to algorithm training. Quality is divided into five dimensions: basic imaging quality, face usability, annotation consistency, distributional fairness, and compliance/security. A standardized framework for indicator measurement, graded admission, closed-loop handling, and training feedback is proposed, together with implementable scoring, thresholds, and governance procedures. The study argues that portrait data quality standards should not consider only sharpness or resolution; instead, they should build a quality governance system from the perspectives of algorithmic usability, group balance, and responsibility across the full data lifecycle, so that the standards are explainable, auditable, and re-verifiable.

Keywords Portrait Dataset; Image Quality Assessment; Data Annotation; Quantitative Quality Standard; Training Admission

1 Research Background and Standardization Needs for Quantifying Face Recognition Training Data Quality

Face recognition has become a basic capability in computer vision, identity verification, smart terminals, and public-service systems. With the development of deep convolutional networks, Transformer architectures, and multi-task learning methods, algorithmic performance has continuously improved on public benchmarks. However, misidentification, missed recognition, cross-domain degradation, and group differences remain serious in engineering deployment. The causes of these problems are not entirely on the model side. The image quality, identity labels, scene coverage, and compliance boundaries of the training data all affect the separability of the model feature space and the upper bound of generalization. For portrait datasets in particular, low resolution, strong backlighting, overexposure, profile faces, occlusion, compression artifacts, duplicate samples, and incorrect identity labels are repeatedly learned by the model, which may lead to class-center drift, unclear feature boundaries, and imbalanced group performance.

Recent work on real and synthetic face datasets has made quality evaluation more closely connected to model training. A journal benchmark of face-recognition systems compares authentic and synthetic data under privacy, bias, and generalization constraints [1]. Training-data standards should therefore examine coverage, trustworthy identity labels, and recognition performance together rather than treating data preparation as a separate administrative step.

Under these circumstances, it is practically necessary to establish a quantitative standard for the image quality of portrait datasets used in face recognition algorithm training. On the one hand, algorithm R&D teams need to rapidly clean and stratify large volumes of collected data so that the data entering training is qualified. On the other hand, data governance departments need traceable indicator systems to prove that collected data has lawful sources, reliable quality, and controllable bias. This paper mainly studies how to transform portrait image quality into computable, gradable, and auditable training-admission

standards, thereby providing methodological reference for the construction of face recognition training sets.

2 Research Status and Application Boundaries of Portrait Dataset Image Quality Assessment

2.1 Face Recognition Training Increasingly Depends on Data Quality

Deep face recognition models generally use identity-classification or metric-learning objectives to pull samples of the same person closer while separating different identities. Incorrect labels can undermine this structure: research on noise-resilient face recognition distinguishes open-set and closed-set label errors and applies different corrective strategies to them [2]. Quality inspection should therefore measure label consistency before and during training.

Recognition in complex environments also depends on whether the recorded face retains useful identity features. A study of blind face restoration introduced benchmark datasets and task-driven measures tied to facial landmarks and identity similarity [3]. Face-image quality assessment based on perturbed feature matching further defines quality by the stability of information used for recognition [4]. These approaches suggest that image admission should consider task utility as well as visible sharpness.

2.2 Portrait Image Quality Assessment Shifts from Visual Quality to Training Usability

Traditional image-quality assessment emphasizes sharpness, noise, contrast, and compression artifacts. Training usability requires a wider view: a second journal evaluation of authentic and synthetic face data explicitly examines demographic bias, pose change, occlusion, and domain adaptation [5]. Quality thresholds should accordingly be calibrated against the intended recognition task and monitored across these conditions.

Dataset-level balance is also part of recognition usability. A recent demographically structured face dataset supplies balanced groups by age, gender, and ethnicity to support fairness testing in face biometrics [6]. This motivates checking both the quality of individual face crops and the distribution of usable samples across groups and acquisition conditions.

2.3 Fairness, Privacy, and Governability Become New Dimensions of Quality Standards

Portrait data quality has distributional and ethical dimensions. A bias-aware evaluation of facial-attribute datasets documents imbalances across demographic categories and illustrates why aggregate accuracy can conceal uneven coverage [7]. If a dataset is highly concentrated in one population or acquisition scenario, good single-image scores alone cannot establish stable performance for underrepresented groups.

Faces are sensitive biometric data, so training-set admissibility also depends on provenance and permitted reuse. Research on privacy-preserving generation of face-recognition data demonstrates an alternative that reduces reliance on directly collected facial images while retaining a training objective [8]. A quality process should record permissions, access, retention, and deletion alongside image and annotation scores.

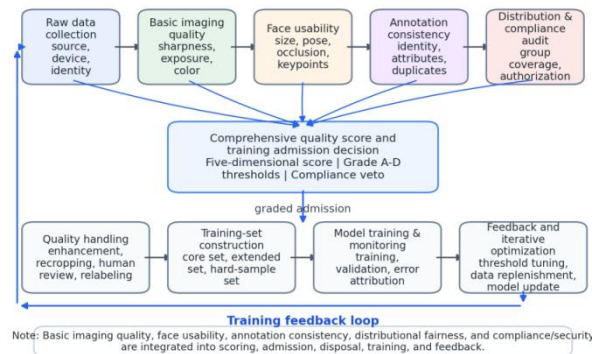


Fig. 1. Portrait dataset quality quantification and training admission process

Fig.1 shows the full process of portrait training data from raw collection to basic imaging-quality detection, face-usability assessment, annotation-consistency verification, distribution and compliance review, comprehensive scoring, graded admission, and training feedback closure. Compared with a static framework, this figure highlights the connection between quality assessment and training management, clearly presenting a standardized governance path of "detection-scoring-handling-training-feedback".

3 Key Problems Caused by the Absence of Standards for Portrait Training Data Quality

3.1 A Single Image Indicator Cannot Represent Training Value

Current portrait dataset construction often filters samples by resolution, detection confidence, or subjective manual judgment. This approach is simple, but it easily ignores the interaction among multiple quality dimensions. A high-resolution image may contain motion blur or strong occlusion. An image with high detection confidence may harm training if its identity label is wrong. An image that appears subjectively acceptable may fail to satisfy algorithm-training requirements in pose angle, inter-eye distance, or compression artifacts. Therefore, a single quality indicator can reflect only a local risk and cannot be directly transformed into a training-admission standard.

3.2 Threshold Rules Lack Task Adaptation and Interpretability

Different models, application fields, and training stages have different quality requirements. Pretraining emphasizes scale and diversity and can tolerate a certain amount of noise. Supervised fine-tuning requires accurate labels and scene matching. Validation sets and test sets require stable distributions, confirmed labels, and balanced sampling. If the same threshold is used for all data, valuable difficult samples may be excessively removed, while low-quality redundant samples may enter training. Quality standards should therefore distinguish basic admission thresholds, priority-training thresholds, and reinspection thresholds, and should explain the training effect represented by each threshold.

3.3 Annotation Errors, Duplicate Samples, and Group Bias are Difficult to Expose in Time

Portrait datasets often contain cases such as different people with the same name, the same person under different names, near-duplicate photographs, batch cropping errors, and cross-device color differences. These problems may not be discovered by single-image quality detection, but they have long-term effects on the model. Identity-label errors cause class-center mixing, duplicate samples increase the weight of a certain pose or device, and group-distribution imbalance increases the risk of misrecognition for minority groups. Without establishing dataset-level statistics, performance problems arising during training are difficult to trace back to concrete data causes.

3.4 Compliance Quality is Disconnected from Technical Quality

Many dataset quality-inspection processes emphasize whether images are clear and whether annotations are correct, but pay insufficient attention to authorization evidence, use boundaries, sensitive-attribute handling, and sample withdrawal. In face recognition training, compliance defects mean that even data with high technical quality cannot be used for training or sharing. The absence of compliance quality also prevents data reuse, external model evaluation, and third-party audits, thereby increasing system-governance costs.

Table 1. Multi-dimensional quality indicator system for portrait images

Quality dimension	Key indicators	Calculation or judgment logic	Recommended threshold	Training impact
Basic imaging quality	Sharpness, noise, exposure, contrast	Laplacian variance, mean luminance, noise estimation, compression-artifact detection	Basic score not lower than 70	Affects texture stability
Face usability quality	Face size, pose, occlusion, landmarks	Detection-box ratio, yaw angle, pitch angle, visibility rate of eyes, nose, and mouth	Visibility of key regions not lower than 80%	Affects identity-feature learning
Alignment and cropping quality	Boundary padding, inter-eye distance, rotation angle	Geometric consistency between face box and landmarks; crop-center offset	Center offset not exceeding 10%	Affects batch-training consistency
Annotation consistency quality	Identity labels, attribute labels, duplicate samples	Multi-annotator agreement, near-duplicate retrieval, cross-batch identity verification	Agreement not lower than 98%	Affects class-center purity
Distributional fairness quality	Age, gender, scenario, device	Group proportion, scenario entropy, device-source distribution, long-tail sample ratio	Balanced coverage of core groups	Affects generalization and bias
Compliance/security quality	Authorization, purpose, desensitization, traceability	Authorization-file verification, access rights, processing logs, withdrawal mechanism	End-to-end traceability	Affects usable boundaries

Table 1 decomposes portrait image quality into six quantifiable indicators and avoids judging data usability only by resolution or manual visual perception. The thresholds are minimum training requirements, and each item should be determined according to the model structure, acquisition equipment, application risk, and other factors. The core value of this indicator system is to unify single-image quality, identity-label quality, and the overall structure of the dataset within one acceptance framework.

4 Construction Strategy for Multi-dimensional Quantitative Standards of Portrait Images for Algorithm Training

4.1 Construct a Five-Dimensional Quality Scoring Model

This paper argues that portrait dataset quality can be divided into five first-level dimensions: basic imaging quality, identity usability, annotation consistency, distributional fairness, and compliance/security. Basic imaging quality assesses whether the image contains sufficient visual information. Identity usability assesses whether key facial regions can support identity-feature extraction. Annotation consistency assesses whether identities, attributes, and sample boundaries are reliable. Distributional fairness assesses whether the dataset contains major groups and scenarios. Compliance/security assesses whether the data has lawful sources and auditable links.

The comprehensive quality score can be expressed by weighted summation: $Q = w_1q_1 + w_2q_2 + w_3q_3 + w_4q_4 + w_5q_5$. Here, q_1 to q_5 denote the normalized scores of the five first-level dimensions, and w_1 to w_5 denote their weights. For ordinary training sets, the weights of basic imaging quality and identity usability can be emphasized. For high-risk applications such as finance, government affairs, and public services, the weights of annotation consistency, distributional fairness, and compliance/security should be increased. Weight determination should be inversely calibrated by validation-set false-recognition rate, missed-recognition rate, and group-difference indicators, rather than relying solely on experience.

4.2 Establish Graded Admission and Differentiated Handling Rules

After quality scoring, graded admission should be implemented according to training use. Grade A samples can enter the core training set. Grade B samples can enter the extended training set or hard-sample set. Grade C samples should enter an enhancement/repair and manual-review queue. Grade D samples should be removed or used only for training quality-enhancement models. This grading mechanism can preserve the value of difficult samples while preventing low-quality data from contaminating identity-feature learning. Problems such as occlusion, low light, and low resolution may be repaired through enhancement, super-resolution, denoising, or recropping; however, image enhancement cannot be used as a direct basis for automatic release during reinspection.

Table 2. Admission grading and handling rules for portrait datasets

Grade	Comprehensive score interval	Recommended use	Handling action	Reinspection requirement
Grade A	85 to 100	Core training and validation sampling	Direct admission and retention of audit records	Quarterly sampling inspection
Grade B	70 to 84	Extended training and hard-sample training	Down-weighted use or stratified sampling by scenario	Batch reinspection
Grade C	55 to 69	Enhancement/repair and manual verification	Relighting, super-resolution, recropping, or relabeling	Admit only after passing reinspection
Grade D	Below 55	Not used for identity training	Remove, isolate, or use only for quality models	Must not be directly used for training
Compliance anomaly	Any score	Suspend use	Verify authorization, delete, or anonymize	Reassess after compliance passes

Table 2 converts the comprehensive score into data-admission actions, so that quality assessment can directly serve training management. The focus of this grading rule is not simply to delete low-score samples, but to distinguish technical-quality problems that can be repaired from annotation or compliance problems that cannot be released. Stratified sampling and reinspection improve training-set purity while preserving some hard-to-recognize but valuable samples.

4.3 Form a Closed Loop of Indicator Measurement, Human Review, and Training Feedback

Standardized quality assessment includes three parts: automatic measurement, human review, and training feedback. Automatic measurement mainly covers sharpness, exposure, face boxes, landmarks, pose, duplicate samples, and distribution statistics. Human review mainly handles boundary samples, identity conflicts, occlusion disputes, and compliance materials. Training feedback pushes the error samples found on the validation set back into the quality-inspection system to analyze whether the error was caused by low-quality images, label noise, group imbalance, or scenario migration. The closed-loop mechanism transforms data-quality standards from static thresholds into dynamic optimization rules.

In engineering implementation, data sample cards and data batch cards can be created. A sample card records the image source, acquisition device, quality score, identity label, annotator, review status, and authorization status. A batch card records acquisition time, scenario distribution, group coverage, duplicate rate, low-quality rate, reinspection rate, and final admission rate. This approach meets model-training needs and also facilitates subsequent audit and responsibility tracing.

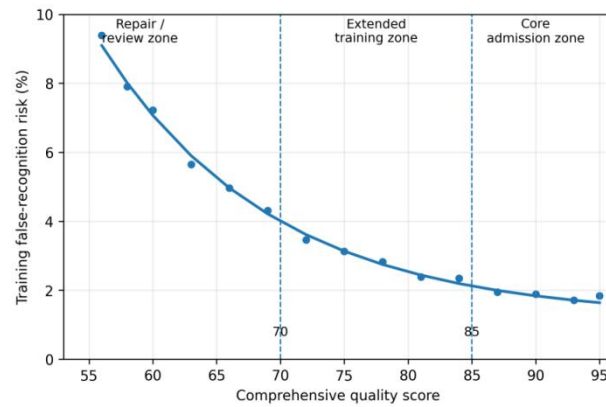


Fig. 2. Relationship between comprehensive quality score and training false-recognition risk

Fig.2 uses example statistics from standardized quality-inspection batches to show the negative relationship between comprehensive quality score and false-recognition risk. Samples with scores below 70 generally have higher misjudgment risk and should be repaired or rechecked. Samples with scores above 85 are more suitable as core training data. In the figure, the comprehensive quality score is both a management indicator and a basis for training-sampling weights and error tracing.

4.4 Improve Privacy Protection and the Minimum-Necessity Principle

Compliance/security quality in portrait datasets should be a one-vote veto item. Even if a sample is clear, accurately labeled, and evenly distributed, it cannot enter the training set if authorization is not obtained, the use exceeds the authorized scope, or the collection source cannot be proved. Specific strategies include determining the purpose and retention period before collection, collecting only portrait regions and auxiliary attributes necessary for training, masking nonessential backgrounds, identity documents, license plates, bystanders, and other sensitive information, controlling access permissions for trained output models, and establishing a locatable deletion mechanism for samples whose authorization has been withdrawn.

Quality standards should not use sensitive attributes themselves as a reason for arbitrary expansion of collection. Attributes such as age, gender, and skin tone may be used for fairness assessment and bias detection, but they should be handled under minimum necessity, graded authorization, and secure isolation. When authorization for sensitive attributes cannot be obtained, non-sensitive proxy indicators such as scenario, device, lighting, and region may be used for auxiliary analysis, but the limitations of these proxy indicators must be explained in the report.

4.5 Embed Quality Scores into Training Strategies

The ultimate purpose of quality quantification is not to generate static reports, but to improve the training process. First, quality scores can be used as sampling weights. In the early stage of training, high-quality samples with stable labels can occupy a larger proportion, while low-quality but valuable hard samples can be introduced later through curriculum learning. Second, quality scores can be used to adjust the loss function, reducing the weight of samples suspected of label noise to mitigate class-center contamination. Third, quality scores can be used to build validation sets by separately creating ideal-quality, regular-quality, and hard-quality validation subsets, so as to differentiate model performance under standard and complex scenarios. Fourth, data recollection lists and model-improvement lists can be determined according to the quality dimensions of error samples.

Different datasets should have different quality standards. Identity-verification training sets based on ID-style photographs should raise the thresholds for frontal face, lighting, cropping, and label consistency. Access-control, station, and mobile-terminal training sets should increase coverage for pose, occlusion, device differences, and low-light scenarios. Algorithmic fairness evaluation sets should raise indicators for group coverage and sample balance. Privacy-protection research datasets should establish specialized evaluation standards among identifiability, attribute preservation, and visual quality.

5 Application Conclusions and Extension Value of Face Recognition Training-data Quality Governance

The quality standard for portrait datasets used in face recognition algorithm training should shift from single-image sharpness evaluation to a multi-dimensional, task-oriented, and auditable comprehensive quantitative system. Using basic imaging quality, identity usability, annotation consistency, distributional fairness, and compliance/security as the standard framework, this paper proposes an indicator system, admission grading, handling rules, and a training-feedback mechanism. The study finds that only by placing image quality, label quality, distribution quality, and compliance quality on the same governance chain can training-set noise be effectively reduced, model generalization be improved, and a data foundation be provided for the trustworthy application of face recognition technology.

Future work can proceed in three directions. First, public portrait-quality evaluation benchmarks across devices, scenarios, and populations should be created so that the relationship between quality scores and recognition performance can be verified more stably. Second, the adaptive coupling mechanism between quality scoring and model-training losses should be explored, allowing data quality to directly participate in training optimization. Third, the systems for data withdrawal, de-identification, and third-party auditing of sensitive biometric information should be completed, so that technical quality and social-governance needs can be unified.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Melzi, P., Tolosana, R., Vera-Rodriguez, R., Kim, M., Rathgeb, C., Liu, X., DeAndres-Tame, I., Morales, A., Fierrez, J., Ortega-Garcia, J., Zhao, W., Zhu, X., Yan, Z., Zhang, X.-Y., Wu, J., Lei, Z., Tripathi, S., Kothari, M., Zama, M. H., ... Marras, M. (2024). FRCSyn-onGoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. *Information Fusion*, 107, 102322.
2. Wang, L., Gong, X., Zhang, J., & Chen, R. (2025). Anti-noise face: A resilient model for face recognition with labeled noise data. *Signal Processing: Image Communication*, 134, 117269.
3. Zhang, P., Zhang, K., Luo, W., Li, C., & Wang, G. (2024). Blind face restoration: Benchmark datasets and a baseline model. *Neurocomputing*, 574, 127271.
4. Zhao, Y., Wang, M., Hu, J., Deng, W., & Li, C.-G. (2025). A perturbed match filtering approach for face image quality assessment. *Pattern Recognition*, 164, 111492.
5. DeAndres-Tame, I., Tolosana, R., Melzi, P., Vera-Rodriguez, R., Kim, M., Rathgeb, C., Liu, X., Gomez, L. F., Morales, A., Fierrez, J., Ortega-Garcia, J., Zhong, Z., Huang, Y., Mi, Y., Ding, S., Zhou, S., He, S., Fu, L., Cong, H., ... Menotti, D. (2025). Second FRCSyn-onGoing: Winning solutions and post-challenge analysis to improve face recognition with synthetic data. *Information Fusion*, 120, 103099.
6. Sufian, A., Ghosh, A., Barman, D., Distant, C., Leo, M., & Sultana, F. (2026). DemoFace: A demographic pixelated face biometric image-text embedding dataset with fairness baselines. *Pattern Recognition*, 175, 113091.
7. Cascone, L., Nappi, M., Pero, C., & Wang, X. (2026). A framework for bias-aware dataset evaluation in soft facial attribute recognition. *Pattern Recognition*, 172, 112416.
8. Tomašević, D., Boutros, F., Damer, N., Peer, P., & Štruc, V. (2024). Generating bimodal privacy-preserving data for face recognition. *Engineering Applications of Artificial Intelligence*, 133, 108495.

Biographies

1. **Yitong Fu** Bachelor of Information Technology from the University of Wollongong, Australia; currently pursuing a Master's degree in Secondary Education (Computing) at Macquarie University. Her research interests focus on computer education and the application.

面向人臉識別算法訓練的人像數據集圖像多維度質量量化標準研究

付翊彤¹

¹麥考瑞大學，悉尼，澳大利亞，2109

摘要：人臉識別演算法訓練效果的好壞，既和模型結構、損失函數有關，也和人像數據集圖像品質、標註品質、分佈結構、合規治理水準有關。為瞭解決目前人像數據採集、訓練準入中品質指標分散、主觀篩選比例高、低質樣本處理無統一標準等問題，本文提出面向演算法訓練的人像圖像多維品質量化標準。將品質分為成像基礎品質、人臉可用品質、標註一致性品質、分佈公平品質、合規安全品質五個方面，提出指標測量、分級準入、處置閉環、訓練回饋的標準化框架，給出可落地的評分、閾值、治理流程。研究認為人像數據品質標準不能只看清晰度或者解析度，應該從演算法可用性、羣體均衡性、數據全生命週期責任三個方面來制定品質治理體系，使之具有可解釋性、可審計性、可複檢性。

關鍵詞：人像數據集；圖像品質評價；數據標註；品質量化標準；訓練準入

1. 付翊彤，澳洲伍倫貢大學資訊科技學士，現為澳洲麥考瑞大學中學教育（電腦）專業碩士研究生。主要研究方向為電腦教育與人工智慧應用。擁有多項國家實用新型及發明專利。