Resource Allocation and Optimization
Volume 1 · Number 1 · 2026

Four References per Person: Pose Coverage as a Gallery-Allocation Choice

Blake Butler[1] · Michael Coleman[2] · Paul Lewis[3]

[1] Serbia · [2] Montenegro · [3] Slovenia

Original publication: October 31, 2026 | Article 702

Abstract

Reference count does not specify reference coverage. We allocated four CMU photographs to each of twenty identities, either uniformly from available candidates or with one photograph from each source pose. Across twenty fixed allocation trials, equal-person query accuracy averaged 91.41% for uniform selection and 95.97% for pose-balanced selection. The paired gain was 4.56 percentage points, with a working interval from 1.88 to 7.24. Individual trial comparisons ranged from a small disadvantage to a larger benefit. Pose coverage helped under this gallery budget, conditional on available annotations, a familiar roster, and a simple whole-image distance rule.

Equal-count reference selection using public source labels. Twenty repeated gallery allocations reuse the same people and queries and do not create independent participant samples.

Four References per Person: Pose Coverage as a Gallery-Allocation Choice

> Equal-count reference selection using public source labels. Twenty repeated gallery allocations reuse the same people and queries and do not create independent participant samples.

Abstract

Reference count does not specify reference coverage. We allocated four CMU photographs to each of twenty identities, either uniformly from available candidates or with one photograph from each source pose. Across twenty fixed allocation trials, equal-person query accuracy averaged 91.41% for uniform selection and 95.97% for pose-balanced selection. The paired gain was 4.56 percentage points, with a working interval from 1.88 to 7.24. Individual trial comparisons ranged from a small disadvantage to a larger benefit. Pose coverage helped under this gallery budget, conditional on available annotations, a familiar roster, and a simple whole-image distance rule.

Spending a fixed reference budget

A gallery containing four photographs of a person can represent four distinct viewpoints or repeatedly show only a few. Both galleries satisfy the same image-count budget. If the eventual queries vary in viewpoint, this allocation decision may matter as much as the nominal number of references. Reference count is therefore an incomplete description of the information supplied to a matching system.

Our research question asks whether explicitly covering four source pose categories improves closed-roster lookup relative to uniform selection from the same candidate photographs. Both policies receive four references per person. The comparison uses source metadata to define one allocation rule and leaves the other blind to pose. Query labels and recognition outcomes never choose which photographs enter either gallery.

Meng et al., 2026 propose stacked multi-view identity mosaics for subject-preserving video generation. Its attention to multiple views motivates a modest precursor question about a reference budget. We do not construct its mosaics or generate videos. The present experiment asks whether view coverage changes a simple downstream identity lookup when image count is held fixed.

Cost-sensitive decision analysis in Elkan, 2001 helps distinguish the resource budget from the performance objective. Our resource is a count of already available photographs, and the outcome is correct roster identity. We do not know the monetary cost of collecting a pose, the burden of annotating it, or the utility of each possible identification error. Those quantities cannot be inferred from a favorable accuracy difference.

Coverage and redundancy are different allocation mechanisms

Uniform selection can produce a useful varied gallery, but it can also select two photographs from one pose while omitting another. Pose-balanced selection prevents that particular redundancy. It does not guarantee diversity in lighting, facial expression, background, or image quality. The policy controls a discrete source label, not every visual dimension that could influence matching.

The covering problem in Gonzalez, 1985 offers a related geometric perspective on representative selection. Our policy is not a farthest-point or k-center algorithm and inherits no approximation guarantee from it. It allocates by annotated categories rather than by measured pairwise image distance. That

distinction is useful because a gallery can be balanced in labels without maximizing coverage in an image metric.

Group-equivariant learning in Cohen and Welling, 2016 and group-based augmentation analysis in Chen et al., 2020 address invariance through mathematically specified transformations. Source photographs labeled left or right do not form an exact transformation group of otherwise identical images. We use these works to distinguish representation invariance from reference coverage, with no claim that category balance supplies their theoretical guarantees.

Early recognition work in Samaria and Harter, 1994, Turk and Pentland, 1991, and Belhumeur et al., 1997 illustrates other ways of using multiple observations or learned representations. None is fitted in this experiment. Our nearest-gallery rule keeps the allocation operation visible: changing the chosen reference set changes which original images compete to identify each fixed query.

A candidate pool that supports both policies

The source is Mitchell, 1997, with its original educational description and filename conventions in Carnegie Mellon University, 2026. The archive includes twenty people captured under different expressions, head poses, and eye conditions. We retain the usable quarter-size grayscale images and exclude the source-marked bad captures. No new image acquisition or participant attribute is inferred.

Gallery candidates are the 158 open-eye captures with neutral or happy expressions, across four source pose categories. Most people contribute eight candidates; two supply seven because of source glitches. Every person still has at least one candidate in each pose, so both four-reference policies are feasible for all twenty identities without imputing an image.

Queries are the 155 open-eye sad or angry captures across the available poses. Their identities are all present in the candidate gallery. Expression separation prevents a query capture from entering the reference set, but it does not create an unseen-person evaluation. The same query set is reused for every trial and both policies.

The four source categories are left, right, straight, and up. They are accepted as filename labels rather than independently estimated angles. Balanced selection consequently means one photograph from each category. It does not mean uniform angular spacing, equal three-dimensional visibility, or a geometrically optimal mosaic of the person's appearance.

Dataset variation in Cao et al., 2017 makes clear why coverage can involve many more dimensions than these four labels. Learned representations in Schroff et al., 2015 may also respond differently to gallery variation than our baseline. The experiment does not establish that the same allocation gain would occur with another dataset, a detector-and-alignment pipeline, or a neural identity embedding.

Twenty fixed allocations at the same count

For each person and each of twenty trial labels, a deterministic hash-derived seed controls reference selection. The uniform policy chooses four distinct candidates without replacement. The balanced policy independently selects one available candidate within each of the four poses. Where a pose has only one usable candidate, that candidate is necessarily selected by the balanced rule.

Each policy therefore uses exactly eighty photographs per trial, four for every identity. The chosen capture identifiers are stored for all forty policy-trial menus. No trial is discarded because one policy performed poorly. The seed and menu rules are fixed before computing any query recognition outcomes.

Within a trial, a query is compared against all eighty selected images using raw quarter-size pixels divided by 255. Squared Euclidean distance supplies the nearest reference, and its identity is returned. Exact ties follow the fixed capture ordering. There is no image centering, face crop, fitted projection, quality filtering, or adjustment based on the query's pose.

All candidates compete in one gallery. Improving the coverage of the correct identity can help, while improving competing identities' coverage can also make some queries harder. Thus, the policy's advantage is not guaranteed simply because every source pose appears. The net recognition difference is an empirical consequence of the full gallery and fixed query distribution.

Quality and identity conditioning remain separate

Kim et al., 2022 and Meng et al., 2021 consider quality-related aspects of identity representation. Our allocation policies do not read a quality estimate or choose the sharpest available reference. A balanced menu may include a photograph that a quality-aware selector would avoid. The observed gain cannot identify whether pose coverage outweighs quality selection in another matching system.

Reference-conditioned generation in Li et al., 2023a, Wang et al., 2024, and Guo et al., 2024 uses identity information through different model designs. Kumari et al., 2022 study customization of generation. Their relevance is the broader importance of how subject evidence is supplied, rather than proof that a balanced four-photo gallery will improve their outputs. No forward generation or model-specific reference encoding is executed here.

The policy also assumes gallery pose metadata are available. Uniform selection needs only the candidate identities, whereas balanced selection additionally consumes category labels. Those annotations have an informational cost even if the photograph count is identical. We report that asymmetry directly instead of describing the comparison as equal in every conceivable resource.

What the repeated trials contribute to inference

We record correctness for every query, policy, and trial. For each identity, we then average across that person's observed queries and the twenty allocation trials. The regression outcome is the balanced-policy mean minus the uniform-policy mean for that identity. This produces twenty person-level paired rows, not thousands of independent query-trial observations.

An intercept-only OLS model estimates the equal-person gain with HC3 covariance and t19 intervals. The rationale for heteroskedasticity-robust uncertainty comes from White, 1980. There are no controls or fixed effects, and R-squared is zero because the model summarizes a mean. The interval conditions on the fixed source people, candidate pool, and finite set of generated menus.

Dependence-aware validation in Roberts et al., 2017 motivates this aggregation. Repeating gallery selection does not create new people, new query images, or independent training datasets. The displayed interval also does not fully propagate uncertainty over every possible random menu. A larger simulation study could estimate that component more directly without changing the number of source identities available for external generalization.

The weighting discussion in Solon et al., 2015 clarifies why we average people equally. Some source identities have fewer queries because of camera glitches, so pooled accuracy and the equal-person estimand differ slightly. Neither is silently substituted for the other. The query-level outputs remain available to evaluate a different, explicitly declared weighting objective.

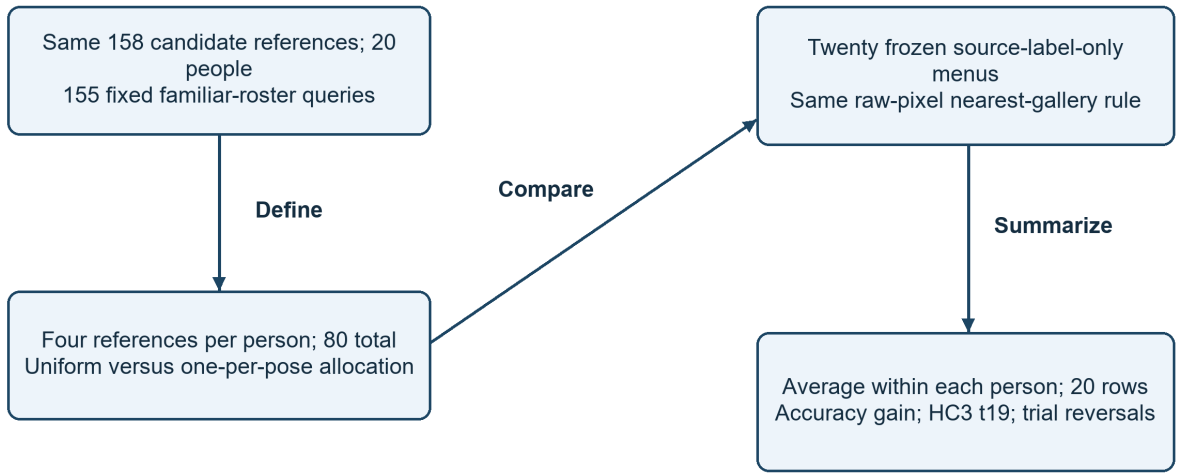


Figure 1. *A pose-aware allocation rule and a pose-blind uniform rule share the same per-person reference count. Source labels enter selection; query outcomes do not.*

Average gains with occasional trial reversals

Uniform selection achieved an equal-person mean accuracy of 0.914092 over the twenty trials. Pose-balanced selection achieved 0.959688. Their paired difference was 0.045595, with standard error 0.012823 and working interval 0.018757 to 0.072433. The favorable mean comparison therefore concerns about 4.56 percentage points of source-roster accuracy under the stated budget.

Table 1. *Pose-balanced minus uniform gallery accuracy, averaged over twenty fixed allocation trials within each person. N equals twenty identities; intercept-only OLS with HC3 covariance and t19 interval. No controls or fixed effects. Both policies use four references per person and the same queries.*

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	0.045595 (0.012823)	[0.018757, 0.072433]	20
M1: fit	R-squared = 0.0000	-	-	20

Model labels: M1 = Pose-balanced minus uniform gallery accuracy.

Fourteen person means favored balanced selection, two favored uniform selection, and four had no average difference. Across the repeated trial-query records, balanced selection produced 2,972 successes and uniform selection 2,833 out of 3,100 opportunities each. Those opportunities reuse the same 155 photographs twenty times, so they are reported as execution counts rather than an independent test-set size.

Individual trial comparisons were less uniform than the average suggests. The pooled-query balanced-minus-uniform difference ranged from about -0.019355 to 0.096774. A randomly instantiated balanced gallery could therefore perform slightly worse than its paired uniform gallery in this finite set of menus. The average benefit is not a guarantee for every permitted allocation realization.

The reference budget does not settle the deployment decision

The result supports the proposition that deliberate coverage can be more useful than redundant references under a fixed photo count. It does not establish that four references are optimal, that another pose taxonomy would perform worse, or that the annotation step pays for itself. Those decisions would require additional budgets, costs, and query distributions that were not evaluated.

Coverage is valuable relative to a destination distribution

The source queries contain several head poses for each person. Covering those categories is therefore a plausible strategy for this particular destination, even though the allocation rule never reads the query outcomes. A deployment dominated by one pose could have a different objective. Reference categories useful for broad coverage might then consume budget that a more concentrated policy would spend on quality or expression variation within the frequent view. That possibility is a design boundary, not a result measured by the current query set.

The uniform policy is also a meaningful competitor rather than a guarantee of poor coverage. Some random menus already contain all four poses, while others omit one or more. Balanced selection removes that source of variation but can still choose a less useful photograph within a pose. The negative end of the observed trial-gap range is consistent with this limited guarantee: category coverage alone does not impose a mathematical ordering on nearest-gallery accuracy.

Feasibility matters before such a policy is reused. The current source candidate pool happens to contain every required pose for every person despite a few camera glitches. If an identity lacked a category entirely, one-per-pose allocation would require a new acquisition, a relaxed rule, or a different budget. We have not manufactured a reference to cover an absent category. The comparison therefore assesses selection among available images, leaving the cost and risk of creating missing reference information unresolved.

Finally, repeated allocation tests address variability in which existing references are chosen. They cannot reveal the effect of acquiring genuinely new pictures or photographing the same people under a new session. A reference-selection gain and an acquisition gain are different empirical objects. The saved menus make the former reproducible while preserving a clear boundary around the latter.

Williams et al., 2009 distinguish high-level workload dimensions from actual hardware limits. Kalibera and Jones, 2013 likewise emphasize measurements appropriate to a computational performance claim. We do not infer latency, storage, energy, or acquisition effort from using eighty references. Both policies have the same count, but their annotation requirements and any real-world collection costs remain outside the measured resource.

The measurement-error perspective in Carroll et al., 2006 keeps whole-image lookup separate from latent facial identity strength. Stevens, 1946 requires the performance unit to remain an accuracy fraction, not a percentage of identity preserved in a generated video. Because the baseline uses complete frames, background and clothing can support both apparent success and policy differences.

Selection concerns in Varma and Simon, 2006 and leakage analysis in Kapoor and Narayanan, 2023 motivate keeping query outcomes away from gallery construction. That boundary is enforced here. It does not make the secondary study preregistered or certify scientific novelty. The two policies and twenty trial labels are analyst-specified tools for a transparent source-data comparison.

Influence considerations in Cook, 1977 motivate preserving the person-level contrasts, including unfavorable and unchanged cases. We do not remove an identity to strengthen the mean gain.

Documentation principles in Gebru et al., 2021 and reuse principles in Wilkinson et al., 2016 support keeping exact menus, source availability, trial outcomes, and weighting rules with the reported effect.

Conclusion

At four references per person, deliberate coverage of the four source poses improved average lookup in this small familiar-roster experiment. The same photo count nevertheless allowed trial-level reversals and did not account for annotation or collection costs. The defensible lesson is to specify how a reference budget is allocated, alongside how large it is. Evidence about a generative identity system would still require executing that system and assessing its own outputs under the relevant reference policies.

References

1. Roberts, David R., et al. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.” *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.
2. Chen, Shuxiao; Dobriban, Edgar; Lee, Jane H.. (2020). A Group-Theoretic Framework for Data Augmentation. Journal of Machine Learning Research21(245):1-71. https://jmlr.org/papers/v21/20-163.html.
3. Schroff, Florian, et al. "FaceNet: A Unified Embedding for Face Recognition and Clustering" *arXiv preprint arXiv:1503.03832*, 2015. https://arxiv.org/abs/1503.03832.
4. Turk, M.A.; Pentland, A.P.. (1991). Face recognition using eigenfaces. Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.1991.139758.
5. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
6. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.
7. Samaria, F.S.; Harter, A.C.. (1994). Parameterisation of a stochastic model for human face identification. Proceedings of 1994 IEEE Workshop on Applications of Computer Vision. https://doi.org/10.1109/acv.1994.341300.
8. Guo, Zinan, et al. "PuLID: Pure and Lightning ID Customization via Contrastive Alignment." *arXiv preprint arXiv:2404.16022*, 2024. https://arxiv.org/abs/2404.16022.
9. Mitchell, Tom. (1997). CMU Face Images [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5JC79.
10. Kumari, Nupur; Zhang, Bingliang; Zhang, Richard; Shechtman, Eli; Zhu, Jun-Yan. (2022). Multi-Concept Customization of Text-to-Image Diffusion. arXiv:2212.04488. https://arxiv.org/abs/2212.04488.
11. Cohen, Taco S.; Welling, Max. (2016). Group Equivariant Convolutional Networks. Proceedings of the 33rd International Conference on Machine Learning. PMLR48:2990-2999. https://proceedings.mlr.press/v48/cohenc16.html.
12. Carnegie Mellon University. (2026). Neural Networks for Face Recognition: companion data and file documentation [Original educational archive; accessed8November2026]. https://www.cs.cmu.edu/~tom/faces.html.
13. Meng, Zijie; Liu, Jiwen; Liu, Yufei; Tong, Chengzhuo; Liu, Xiaoqiang; Zhang, Yuanxing; Xu, Yulong; Wan, Pengfei. (2026). ARGUS: Stacked Multi-View Identity Mosaic Injection for Subject-Preserving Video Generation. arXiv:2606.11670. https://arxiv.org/abs/2606.11670.
14. Li, Zhen, et al. "PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding." *arXiv preprint arXiv:2312.04461*, 2023. https://arxiv.org/abs/2312.04461.
15. Kalibera, Tomas; Jones, Richard. (2013). Rigorous benchmarking in reasonable time. Proceedings of the 2013 international symposium on memory management. https://doi.org/10.1145/2464157.2464160.
16. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.
17. Williams, Samuel; Waterman, Andrew; Patterson, David. (2009). Roofline: an insightful visual performance model for multicore architectures. Communications of the ACM. https://doi.org/10.1145/1498765.1498785.
18. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.
19. Meng, Qiang, et al. "MagFace: A Universal Representation for Face Recognition and Quality Assessment." *arXiv preprint arXiv:2103.06627*, 2021. https://arxiv.org/abs/2103.06627.

20. Kapoor, Sayash; Narayanan, Arvind. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns. https://doi.org/10.1016/j.patter.2023.100804.

21. Cao, Qiong, et al. "VGGFace2: A dataset for recognising faces across pose and age" *arXiv preprint arXiv:1710.08092*, 2017. https://arxiv.org/abs/1710.08092.

22. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

23. Stevens, S. S.. (1946). On the Theory of Scales of Measurement. Science. https://doi.org/10.1126/science.103.2684.677.

24. Wang, Qixun, et al. "InstantID: Zero-shot Identity-Preserving Generation in Seconds." *arXiv preprint arXiv:2401.07519*, 2024. https://arxiv.org/abs/2401.07519.

25. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; ’t Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

26. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J.. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/34.598228.

27. Solon, Gary; Haider, Steven J.; Wooldridge, Jeffrey M.. (2015). What Are We Weighting For?. Journal of Human Resources. https://doi.org/10.3368/jhr.50.2.301.

28. Elkan, Charles. “The Foundations of Cost-Sensitive Learning.” *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 2001, pp. 973-978. https://cseweb.ucsd.edu/~elkan/rescale.pdf.

29. Gonzalez, Teofilo F.. (1985). Clustering to minimize the maximum intercluster distance. Theoretical Computer Science38:293-306. https://doi.org/10.1016/0304-3975(85)90224-5.

30. Kim, Minchul, et al. "AdaFace: Quality Adaptive Margin for Face Recognition." *arXiv preprint arXiv:2204.00964*, 2022. https://arxiv.org/abs/2204.00964.