Population Variation and Equity
Volume 1 · Number 1 · 2026

Equal Pixel Loss, Different Component Counts: A Controlled Glyph-Erasure Experiment

Aaron Scott[1] · Daniel Cooper[2] · Benjamin Lewis[3]

[1] Netherlands · [2] Denmark · [3] Norway

Original publication: November 3, 2026 | Article 598

Abstract

Removing the same number of foreground pixels guarantees equal binary pixel error, but does not guarantee the same connected-component response. We tested this distinction on one image from each of 964 Omniglot background classes. For every glyph, a random policy and a dense-neighborhood policy removed five percent of its foreground pixels. Random erasure increased component count by 0.255 on average under four-neighbor connectivity, but only 0.002 under eight-neighbor connectivity; dense-neighborhood erasure changed neither count in this sample. The large difference between connectivity conventions is itself consequential. The experiment shows that a pixel-loss tie can conceal representation-sensitive structural differences, without establishing legibility, stroke preservation, or the quality of any restoration model.

Deterministic and seeded corruption of public character images. No inpainting network, readability panel, or cultural-heritage restoration system was tested.

Equal Pixel Loss, Different Component Counts: A Controlled Glyph-Erasure Experiment

Deterministic and seeded corruption of public character images. No inpainting network, readability panel, or cultural-heritage restoration system was tested.

Abstract

Removing the same number of foreground pixels guarantees equal binary pixel error, but does not guarantee the same connected-component response. We tested this distinction on one image from each of 964 Omniglot background classes. For every glyph, a random policy and a dense-neighborhood policy removed five percent of its foreground pixels. Random erasure increased component count by 0.255 on average under four-neighbor connectivity, but only 0.002 under eight-neighbor connectivity; dense-neighborhood erasure changed neither count in this sample. The large difference between connectivity conventions is itself consequential. The experiment shows that a pixel-loss tie can conceal representation-sensitive structural differences, without establishing legibility, stroke preservation, or the quality of any restoration model.

1. A pixel-level tie can be designed in advance

Image-reconstruction evaluation often begins with a numerical distance between a reference and a modified image. That distance can be informative while leaving other properties unconstrained. In a binary character image, changing the same number of foreground pixels yields the same squared pixel error regardless of where those pixels occur. The equality is mathematical, not an empirical hypothesis. What remains empirical is whether different deletion locations change a chosen structural summary differently.

This article asks how two equal-budget erasure policies affect connected-component counts, and whether the answer depends on four- or eight-neighbor connectivity. We use a controlled image transformation rather than a fitted restoration system. The design therefore separates the corruption policy, which we construct, from the structural outcome, which we calculate on the resulting images. It cannot establish that a particular restoration architecture succeeds or fails.

The cultural-heritage reconstruction setting in Meng et al., 2025 motivates the need to distinguish pixel agreement from character structure. Our images are from a different resource and are not oracle-bone rubbings. The experiment is consequently a small measurement study relevant to designing an evaluation, rather than a replication of that model or proof that its outputs have any particular defect. A source-specific restoration claim would require the corresponding images and actual restored outputs.

2. Structural evaluation has several possible objects

Inpainting methods such as Pathak et al., 2016, Liu et al., 2018, and Yu et al., 2018b reconstruct missing image content through learned representations and mask handling. Nazeri et al., 2019 makes edge information particularly relevant to the discussion of structure. More recent alternatives in Suvorov et al., 2021 and Lugmayr et al., 2022 provide other restoration strategies. None is run here. Their relevance is to the distinction between how an image is reconstructed and how the reconstructed result is evaluated.

The image-prior approach in Ulyanov et al., 2017 and state-space visual methods in Guo et al., 2024 and Liu et al., 2024 further illustrate that an architectural family does not determine an evaluation criterion. A model may receive favorable pixel scores and still require additional task-specific checks. Conversely, an

unfavorable value on one structural summary need not establish that a human cannot interpret the character. The empirical endpoint in this article is deliberately specified before discussing those broader judgments.

Perceptual metrics in Zhang et al., 2018 and Wang et al., 2004 address information beyond raw squared pixel differences, but neither is computed in this experiment. The source corpus described by Wang and Deng, 2024 is relevant to future character-specific evaluation, without supplying observations to this analysis. We retain these boundaries because metric names can otherwise give a false impression of a broad benchmark when the executed study contains only binary erasure and component counting.

3. The image sample and its deliberate restriction

We use the official public Omniglot background image archive, in the setting documented by Lake et al., 2019. It contains 964 background character classes spanning thirty alphabets. For every class we select the image whose filename ends in `_01.png`, giving one exemplar per class. This deterministic rule avoids selecting the most visually convenient image, but it is not a random sample of writers or all exemplars available for each character.

Every selected image has dimensions 105 by 105 pixels. We convert grayscale values below 128 to foreground and the remaining values to background. This threshold defines the binary object for the entire experiment. No learned segmentation, skeletonization, stroke labeling, or manual correction is used. Anti-aliased edges and threshold choice can affect component structure, so the result is conditional on this representation rather than a claim about an invariant underlying glyph shape.

The thirty alphabet labels are retained as grouping information. We do not treat all 964 classes as independent cultural or linguistic replications, and we do not infer the true identities of writing systems beyond the released labels. A study of restoration utility would also need to consider provenance, script expertise, and the purpose of reconstruction. Those questions cannot be answered by a binary image experiment on a separate character collection.

4. Two locations for exactly the same erasure budget

For each binary exemplar we count its foreground pixels and set the deletion budget to the integer floor of five percent of that count, with a minimum of one pixel. The random policy samples that many foreground positions without replacement using seed 20261107. There is one random draw per glyph. We do not average over repeated masks or describe the resulting uncertainty as if it included an entire distribution of random corruption patterns.

The dense-neighborhood policy scores each foreground position by the number of foreground pixels in its three-by-three neighborhood. It deletes positions with the highest scores, resolving ties by stable flattened pixel order. This heuristic selects locally dense regions, but does not recover a true stroke center, junction, or skeleton. Its name describes the actual neighborhood calculation. Interpreting it as anatomically or linguistically informed character damage would overstate what the rule knows.

Both policies delete exactly the same foreground count in the same glyph and leave all other pixels unchanged. Hence their binary squared error relative to that glyph is exactly equal by construction. This equality is included as a design check. It is not presented as a surprising experimental result or as evidence that the two damaged glyphs look equally acceptable. The structural measurements are needed precisely because the pixel-budget constraint leaves their spatial arrangement free.

5. Four neighbors and eight neighbors ask different questions

Under four-neighbor connectivity, a foreground pixel connects to foreground pixels immediately above, below, left, and right. Eight-neighbor connectivity additionally permits diagonal connections. We count connected foreground components under each rule before and after erasure. The outcome is the post-erasure component count minus the original count calculated under the same convention. A positive value indicates more disconnected foreground pieces by that definition.

This component count is only one aspect of topology. It does not count holes, encode loop geometry, identify strokes, or establish semantic similarity. In particular, deleting pixels from a dense interior can create holes while leaving the number of foreground components unchanged. A zero component change therefore does not mean the image is undamaged or structurally identical. The experiment's conclusion must retain that distinction even when one policy has zero change throughout the observed sample.

The graph viewpoint in Ortega et al., 2017 provides a useful way to understand the neighborhood convention: foreground positions and adjacency rules define a graph whose connected components we count. Changing the adjacency changes the graph, so a change in the result is not necessarily a software inconsistency. It reveals a measurement choice. A later restoration evaluation should specify which notion of adjacency is appropriate for the intended structure rather than assuming that all component counters measure the same object.

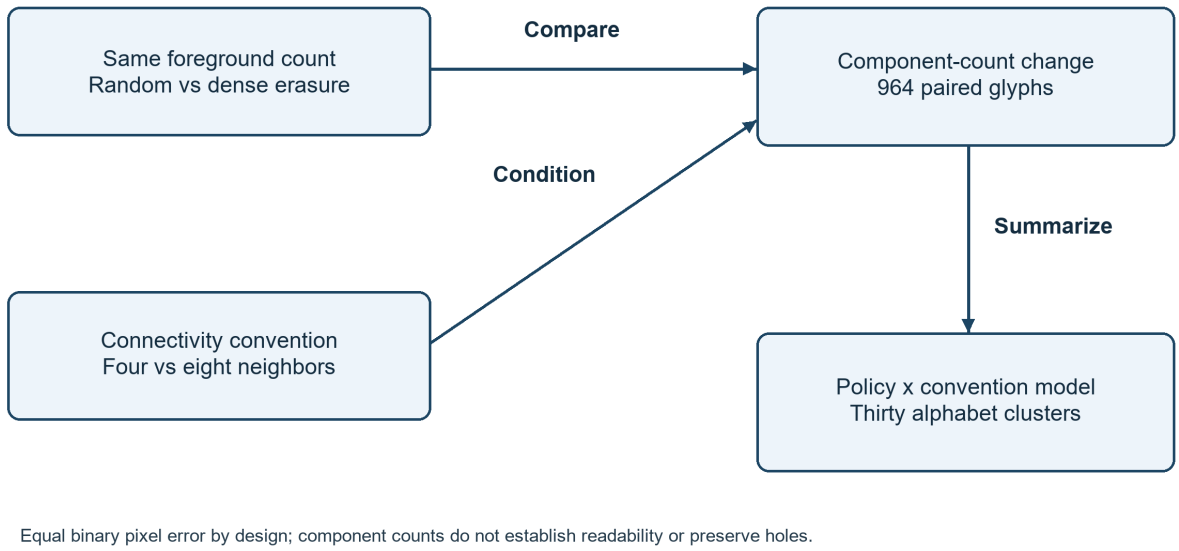


Figure 1. The same glyph and deletion count enter two placement rules, each evaluated under two connectivity conventions.

6. A paired factorial summary

There are four rows per glyph: random erasure with four neighbors, random erasure with eight, dense-neighborhood erasure with four, and dense-neighborhood erasure with eight. The resulting dataset contains 3,856 rows. Ordinary least squares models component-count change using the dense-policy indicator, the eight-neighbor indicator, their interaction, and alphabet fixed effects. The reference condition is random erasure under four-neighbor connectivity.

We cluster covariance by alphabet, allowing the repeated conditions and characters in each alphabet to share errors. The procedure follows the dependence considerations in Cameron and Miller, 2015, uses

thirty clusters, a small-sample correction, and a t reference with twenty-nine degrees of freedom. Heteroskedasticity correction alone, as in White, 1980, would not describe that grouping. The interval does not cover uncertainty from alternative thresholds, different exemplars, or additional random mask draws that were not executed.

Table 1. Policy and connectivity contrasts in the change of foreground component count.

Model	Term	Estimate (SE)	95% CI	N
M1	Dense vs random	-0.2552 (0.0223)	[-0.3009, -0.2095]	3856
M1	Eight vs four neighbors	-0.2531 (0.0222)	[-0.2986, -0.2076]	3856
M1	Placement x convention	0.2531 (0.0222)	[0.2076, 0.2986]	3856
M1: fit	R-squared = 0.1635	-	-	3856

Model labels: M1 = Mask placement by connectivity component-change contrast.

The regression is a compact description of the paired transformations, not a model for independently sampled natural damage. Alphabet fixed effects absorb average differences across released groups. The reported R-squared of 0.1635 describes variation in this constructed outcome. It cannot be interpreted as the proportion of readability explained by the erasure rule, or as a measure of how well a restoration system recovers original character identity.

7. The connectivity convention dominates the random-erasure result

Random erasure produces an average component-count increase of 0.255187 under four-neighbor connectivity. The fraction of glyphs whose count increases is 0.220954. Under eight-neighbor connectivity, the average increase is only 0.002075, with increases in two of the 964 glyphs. The same modified pixels therefore support very different fragmentation summaries depending on whether diagonal links are allowed. This difference directly answers the second research question.

Dense-neighborhood erasure yields zero component-count change for every selected glyph under both conventions. Relative to random four-neighbor erasure, its coefficient is -0.255187, with standard error 0.022339 and interval -0.300876 to -0.209497. The eight-neighbor coefficient is -0.253112, with standard error 0.022232 and interval -0.298581 to -0.207643. The interaction is 0.253112, with interval 0.207643 to 0.298581. Together these coefficients reproduce the much smaller policy difference under diagonal connectivity.

The zero dense-policy change should not be translated into a finding that dense damage is harmless. The rule may remove interior information while leaving connectivity intact. Likewise, an extra component may be a small isolated fragment rather than destruction of a critical character distinction. Without expert reading judgments or a validated recognition task, the component outcome has no established monotonic relationship with comprehension. It measures a geometric property, and the strength of the result lies in keeping that property explicit.

Because the mask budgets match exactly, the pixel-loss tie exists for every glyph, not only on average. Yet the component result is heterogeneous across images under random erasure. This separation is the empirical point: a controlled equality on one endpoint does not determine another endpoint's behavior. It does not establish that connected components are always a better quality measure. Whether they are useful depends on the failure mode a future evaluation intends to detect.

8. Agreement and robustness require their own evidence

Agreement concepts in Bland and Altman, 1986, Lin, 1989, and Koo and Li, 2016 clarify why two metrics should not be declared interchangeable because they are both numerical or respond to the same image transformations. We do not calculate an agreement coefficient or intraclass reliability estimate. Measurement-error reasoning in Carroll et al., 2006 instead reminds us that binary thresholding is part of the observed representation. A different threshold or scan process could alter connectivity before any experimental erasure occurs.

Missing-data perspectives in Little and Rubin, 2002 matter because an artificial deletion of known foreground is not the mechanism of natural rubbing loss, erosion, occlusion, or acquisition noise. Robustness traditions in Huber, 1964, Hampel, 1974, and Rousseeuw, 1984 also do not imply that every robust numerical summary preserves semantic content. The experiment evaluates neither robust estimation nor natural damage recovery. Its controlled corruption supplies a measurement contrast whose external meaning must be established separately.

Influence analysis motivated by Cook, 1977 could identify glyphs contributing most strongly to the four-neighbor difference. Such glyphs are not removed here. Multiplicity issues in Benjamini and Hochberg, 1995 would become relevant if many deletion rates, structural endpoints, and neighborhood rules were searched. The present report retains the declared five-percent budget and two connectivity definitions. Interpretation under Wasserstein and Lazar, 2016 favors presenting these effect magnitudes and scope over turning one interval into a universal judgment about image fidelity.

9. A focused next step for restoration evaluation

A stronger restoration study could collect outputs from specified models, preserve the actual missing-region masks, and test whether a structural diagnostic predicts expert-detected errors. It would need a representation suitable for the script, and separate evidence for readability or recognition. Our controlled result suggests including the adjacency definition in that protocol. It does not select a threshold, choose an optimal metric combination, or establish that a four-neighbor convention is preferable to an eight-neighbor one for every character corpus.

The documentation aims of Gebru et al., 2021 support recording exemplar selection, threshold, deletion count, neighborhood score, tie-breaking, and random seed. Wilkinson et al., 2016 motivates preservation of the raw archive and per-glyph transformations and summaries. The finding remains precise: equal binary pixel loss coexists with different component-count responses, and the size of that difference depends strongly on how connectivity is defined. That is a reproducible measurement lesson, with restoration usefulness still requiring task-specific validation.

The dense-policy tie rule is worth preserving because local foreground density is discrete. Many positions can receive the same neighborhood count, and choosing among them can change the geometry of the erased region. Stable flattened order makes the executed rule reproducible but does not make it rotation invariant or linguistically meaningful. A rotated image can change which tied positions are selected even when its local density distribution is similar. That is another reason to treat the observed zero component change as a result for this exact policy, rather than a theorem about every possible interior-erasure strategy.

Likewise, one exemplar per class keeps class weighting transparent while leaving handwriting variability untested. Additional exemplars could have thinner strokes, different diagonal connections, or different threshold responses. They would add a genuinely useful replication dimension rather than merely more transformed copies of the same selected image. The current alphabet-clustered intervals describe the

observed paired sample; they do not quantify the uncertainty that would arise from drawing another writer's image for each class or evaluating a naturally damaged inscription.

References

1. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.
2. 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.
3. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
4. Nazeri, Kamyar; Ng, Eric; Joseph, Tony; Qureshi, Faisal Z.; Ebrahimi, Mehran. (2019). EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. arXiv:1901.00212. https://arxiv.org/abs/1901.00212.
5. Hampel, Frank R.. (1974). The Influence Curve and its Role in Robust Estimation. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1974.10482962.
6. Koo, Terry K., and Mae Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, pp. 155-163. https://doi.org/10.1016/j.jcm.2016.02.012.
7. Ulyanov, Dmitry; Vedaldi, Andrea; Lempitsky, Victor. (2017). Deep Image Prior. arXiv:1711.10925. https://arxiv.org/abs/1711.10925.
8. Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.
9. Lake, Brenden M.; Salakhutdinov, Ruslan; Tenenbaum, Joshua B.. (2019). The Omniglot challenge: a 3-year progress report. https://arxiv.org/abs/1902.03477.
10. Guo, Hang; Li, Jinmin; Dai, Tao; Ouyang, Zhihao; Ren, Xudong; Xia, Shu-Tao. (2024). MambaIR: A Simple Baseline for Image Restoration with State-Space Model. arXiv:2402.15648. https://arxiv.org/abs/2402.15648.
11. Lin, Lawrence I-Kuei. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. https://doi.org/10.2307/2532051.
12. Huber, Peter J.. (1964). Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics. https://doi.org/10.1214/aoms/1177703732.
13. Rousseeuw, Peter J.. (1984). Least Median of Squares Regression. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1984.10477105.
14. Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
15. Lugmayr, Andreas; Danelljan, Martin; Romero, Andres; Yu, Fisher; Timofte, Radu; Van Gool, Luc. (2022). RePaint: Inpainting using Denoising Diffusion Probabilistic Models. arXiv:2201.09865. https://arxiv.org/abs/2201.09865.
16. Wasserstein, Ronald L.; Lazar, Nicole A.. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician. https://doi.org/10.1080/00031305.2016.1154108.
17. Bland, J. Martin, and Douglas G. Altman. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *The Lancet*, vol. 327, no. 8476, 1986, pp. 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.
18. Zhang, Richard, et al. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric" *arXiv preprint arXiv:1801.03924*, 2018. https://arxiv.org/abs/1801.03924.
19. Meng, Zijie; Zeng, Yuanze; Chang, Xiang; Xu, Tianshuo; Chao, Fei; Cao, Xixin; Shang, Changjing; Shen, Qiang. (2025). Orpaint: a zero-shot inpainting model for oracle bone inscription rubbings with visual mamba block. Science China Information Sciences, 68(8), 189102. https://doi.org/10.1007/s11432-024-4493-4.
20. Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.
21. Liu, Yue; Tian, Yunjie; Zhao, Yuzhong; Yu, Hongtian; Xie, Lingxi; Wang, Yaowei; Ye, Qixiang; Jiao, Jianbin; Liu, Yunfan. (2024). VMamba: Visual State Space Model. arXiv:2401.10166. https://arxiv.org/abs/2401.10166.

22. 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.

23. Pathak, Deepak; Krahenbuhl, Philipp; Donahue, Jeff; Darrell, Trevor; Efros, Alexei A.. (2016). Context Encoders: Feature Learning by Inpainting. arXiv:1604.07379. https://arxiv.org/abs/1604.07379.

24. Yu, Jiahui; Lin, Zhe; Yang, Jimei; Shen, Xiaohui; Lu, Xin; Huang, Thomas. (2018b). Free-Form Image Inpainting with Gated Convolution. arXiv:1806.03589. https://arxiv.org/abs/1806.03589.

25. Liu, Guilin; Reda, Fitsum A.; Shih, Kevin J.; Wang, Ting-Chun; Tao, Andrew; Catanzaro, Bryan. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. arXiv:1804.07723. https://arxiv.org/abs/1804.07723.

26. Wang, Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004, pp. 600-612. https://doi.org/10.1109/TIP.2003.819861.

27. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. "Measurement Error in Nonlinear Models." 2006. https://doi.org/10.1201/9781420010138.

28. Ortega, Antonio; Frossard, Pascal; Kovačević, Jelena; Moura, José M. F.; Vandergheynst, Pierre. (2017). Graph Signal Processing: Overview, Challenges and Applications. arXiv:1712.00468. https://arxiv.org/abs/1712.00468.

29. Suvorov, Roman; Logacheva, Elizaveta; Mashikhin, Anton; Remizova, Anastasia; Ashukha, Arsenii; Silvestrov, Aleksei; Kong, Naejin; Goka, Harshith; Park, Kiwoong; Lempitsky, Victor. (2021). Resolution-robust Large Mask Inpainting with Fourier Convolutions. arXiv:2109.07161. https://arxiv.org/abs/2109.07161.

30. Wang, Mei; Deng, Weihong. (2024). A dataset of oracle characters for benchmarking machine learning algorithms. Scientific Data, 11, 87. https://doi.org/10.1038/s41597-024-02933-w.