Journal of Measurement and Evidence
Volume 1 · Number 1 · 2026

Mirror Consistency Without Retraining: What a Symmetric Completion Rule Actually Guarantees

Oliver Collins[1] · Samuel Marshall[2] · Matthew Fisher[3]

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

Original publication: October 2, 2026 | Article 662

Abstract

A predictor can be made exactly mirror consistent by averaging its ordinary output with a reflected-input output mapped back to the original orientation. We applied this construction to a frozen linear predictor of four hidden digit features. Across 1,797 held images, equal-orientation MSE fell from 0.13967 to 0.10313. Canonical-orientation mean error also fell, although 783 individual images became worse in that orientation. The paired-orientation improvement follows from a squared-loss identity; the canonical result does not. This distinction separates an architectural consistency guarantee from an empirical claim about the orientation that an application actually encounters.

Missing-feature reconstruction on public digit block counts. Reflection is a numerical transformation of the reconstruction task, not an assertion that a reflected glyph preserves its semantic identity.

Mirror Consistency Without Retraining: What a Symmetric Completion Rule Actually Guarantees

> Missing-feature reconstruction on public digit block counts. Reflection is a numerical transformation of the reconstruction task, not an assertion that a reflected glyph preserves its semantic identity.

Abstract

A predictor can be made exactly mirror consistent by averaging its ordinary output with a reflected-input output mapped back to the original orientation. We applied this construction to a frozen linear predictor of four hidden digit features. Across 1,797 held images, equal-orientation MSE fell from 0.13967 to 0.10313. Canonical-orientation mean error also fell, although 783 individual images became worse in that orientation. The paired-orientation improvement follows from a squared-loss identity; the canonical result does not. This distinction separates an architectural consistency guarantee from an empirical claim about the orientation that an application actually encounters.

Reflection poses two questions rather than one

A missing-region predictor can produce different answers when the visible image is reflected and the predicted region is reflected back. That discrepancy is measurable even when both versions contain precisely the same observed information under a permutation of coordinates. Averaging the two answers offers a simple route to consistency, but a consistent answer need not be closer to the concealed values for every image or every operational orientation.

We investigate that distinction using a frozen linear completion rule. The research question is what happens to canonical and reflected reconstruction errors when exact horizontal-reflection consistency is imposed after fitting. We separately evaluate the equally weighted pair of orientations. Keeping these quantities separate is essential because the mathematical guarantee concerns the paired objective, while a source collection can consist mostly of canonical images.

Meng et al., 2025 develop a zero-shot inpainting model for oracle-bone inscription rubbings using a visual Mamba block. No part of that model is executed here. The connection is the interpretation of reconstruction properties: a transformation-consistent rule and a visually or semantically faithful completion are different achievements. Our low-dimensional calculation makes the former inspectable without claiming evidence for readability, archaeological validity or the target model's performance.

Equivariant convolutional networks formalize how transformations can be respected by an architecture (Cohen and Welling, 2016). Group-theoretic analysis of augmentation also connects transformations with averaging over related observations (Chen et al., 2020). We use an elementary two-element reflection construction rather than those training procedures. Its exact arithmetic property is proved for this predictor; its empirical error is measured on the stated held block-count vectors.

The observed features and the fixed hidden block

The Optdigits resource supplies an official fitting partition of 3,823 digit records and a test partition of 1,797 (Alpaydin and Kaynak, 1998). Each record has 64 block-count features on an eight-by-eight grid, with values from zero to sixteen. We divide by sixteen before reconstruction. A feature therefore measures a normalized block count, and its squared error is not automatically equivalent to luminance error in a high-resolution photograph.

Four features in the central two-by-two block are concealed. Their zero-based flattened indices are 27, 28, 35 and 36; the other sixty features remain observed. A horizontal reflection permutes the visible features among themselves and the hidden features among themselves. Consequently, applying the reflected-input route never moves an originally concealed target into the observed predictor set.

The mask's symmetry is a substantive design condition. If a different hidden region were used, the reflected mask would need to be tracked alongside the image to preserve the same information boundary. It would be invalid to reflect the values while keeping an incompatible visibility array and then claim a general transformation guarantee. Our construction is restricted to this declared reflection-invariant central block.

The base predictor reuses a locally frozen affine operator fitted only on the official training rows: the training mean, sixteen principal directions and latent ridge parameter 0.01. This is the same captured operator used in an earlier input-error audit, not an independently fitted replication. The current study changes the prediction rule's symmetry, using the unchanged clean test features and no input-noise realization from that earlier audit.

Singular-value decomposition provides the basis calculation (Golub and Reinsch, 1970), while ridge regularization stabilizes the latent solve from visible coordinates (Hoerl and Kennard, 1970). Neither rank nor penalty is retuned against the reflection outcomes. We record the frozen operator's file hash so that the new sensitivity calculation remains attributable to one specific fitted map rather than to an unspecified family of linear restorers.

How the symmetric answer is constructed

For a canonical input, the first route applies the frozen rule directly to its sixty visible values. The second route reflects the full feature indexing, applies the same rule to the reflected visible values, and maps the four predicted targets back to the canonical orientation. The symmetric answer is the arithmetic mean of those two four-dimensional predictions.

Applying the construction to the reflected input gives the reflected version of that same average. This follows because reflecting twice restores the original indexing. The method is therefore equivariant to this one reflection operation: output coordinates transform with the input. It is not an invariant predictor that returns an unchanged target array regardless of orientation, because the concealed target coordinates also need to be permuted.

The prediction uses no concealed block values while forming either answer. Ground truth is accessed afterward for scoring. It also uses no digit labels in fitting or prediction. Reflection can change how a digit would be read, so we make no classification-label preservation assumption. The target for a reflected vector is simply the correspondingly reflected set of observed source block counts.

Full feature vectors from the fitting partition are available when the principal directions and training mean are estimated. That is a supervised reconstruction setting in the broad sense that the fitting resource includes values at the coordinates later hidden during evaluation. Concealing only the held targets protects the test boundary, but it does not make the baseline zero shot. The training information is the same for the base and symmetric rules, so their contrast isolates post-fitting averaging rather than access to additional complete examples.

This averaging can be viewed as a two-route inference procedure, but a fixed affine predictor permits its coefficients and intercept to be combined in advance. We have not measured the runtime of either implementation. Describing the calculation as involving two predictions does not establish a twofold

deployment cost, and precomputability alone does not establish a speed advantage over a particular software pipeline.

Why the paired loss must improve

After mapping both routes into canonical coordinates, there are two candidate predictions for one target vector. Under squared loss, the error of their average equals the average of their two errors minus one quarter of their squared disagreement. Reflection preserves Euclidean distance, so those two errors correspond exactly to the base rule's canonical and reflected reconstruction errors.

The symmetric rule has identical squared error in the two orientations after accounting for the target-coordinate permutation. Its equal-orientation loss is therefore no greater than the base rule's equal-orientation loss for each original image. The difference equals minus one quarter of the mean squared discrepancy between the two back-aligned base predictions. We verify that identity on the executed outputs rather than infer it from a smaller aggregate mean.

This guarantee does not select the better route for the canonical image. Averaging two candidates can worsen the first even as it improves their average risk. Nor does it promise optimality among all completion rules. The agreement constraint merely combines the specified pair; it does not recover information that both routes fail to infer from the visible feature values.

The result is specific to the chosen squared loss and equal orientation weights. Perceptual and structural criteria can respond differently. Denoising representation learning and missing-image modeling offer other reconstruction objectives and assumptions (Vincent et al., 2008; Pathak et al., 2016). The existence of an exact quadratic identity is not a theorem about whether those methods, human viewers or downstream readers would prefer the averaged completion.

For an unequal orientation mixture, the guaranteed comparison changes. A rule optimized for the canonical direction may rationally retain an asymmetric response if that direction dominates the application. Enforcing equal treatment of an uncommon reflected direction can then sacrifice the more relevant error. The group average does not estimate how frequently either direction occurs. It imposes an explicit prediction constraint, whose practical value depends on an independently specified distribution of inputs.

Empirical comparisons and their units

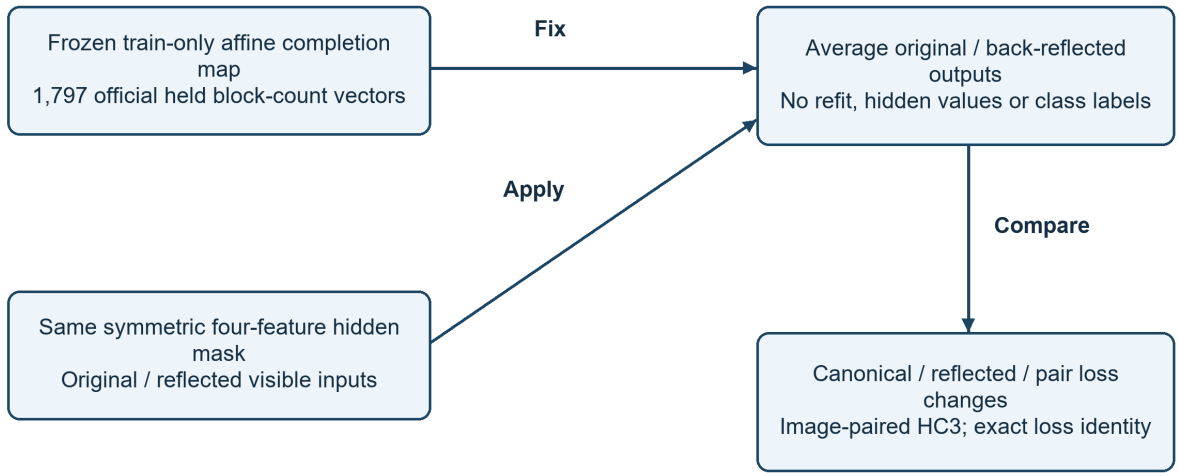

Figure 1. *A canonical prediction and a back-reflected prediction are averaged without refitting. Canonical, reflected and equally weighted pair losses are scored separately; the two orientations are paired views of one source image.*

Each test image supplies three differences: symmetric minus base canonical MSE, symmetric minus base reflected MSE, and their equal-orientation average. Models 1 through 3 estimate the means of those differences using an intercept, HC3 covariance and t1796 intervals. N is 1,797 original images in every model, not 3,594 independent orientations or 7,188 independent hidden features.

HC3 states a working heteroskedastic regression variance (White, 1980). The source documentation describes different training and test writers, but individual test-writer identifiers are not present in these files. We cannot therefore cluster the reported covariance by writer. The intervals also omit fitting-sample uncertainty because they condition on the captured affine map rather than refitting it for different training draws.

Table 1. *Symmetric-minus-base hidden-feature MSE, N=1,797 image pairs, HC3 standard errors and t1796 intervals. Models contain an intercept and no controls; R-squared is zero to numerical precision. The third outcome is a paired mathematical reduction rather than independent evidence of universal robustness.*

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	-0.017177 (0.001953)	[-0.021008, -0.013346]	1797
M1: fit	R-squared = 0.0000	-	-	1797
M2	Mean outcome/contrast	-0.055917 (0.002119)	[-0.060072, -0.051762]	1797
M2: fit	R-squared = 0.0000	-	-	1797
M3	Mean outcome/contrast	-0.036547 (0.000774)	[-0.038064, -0.035029]	1797
M3: fit	R-squared = 0.0000	-	-	1797

Model labels: M1 = Symmetrized minus base canonical-orientation MSE; M2 = Symmetrized minus base reflected-orientation MSE; M3 = Symmetrized minus base equal-orientation-pair MSE.

Before averaging, canonical MSE is 0.120304 and reflected MSE is 0.159044. Their average is 0.139674. Symmetrization gives 0.103127 in either orientation, with orientation equality following from the

construction. We do not present that equality as two separate successful replications; the reflected loss is the same quantity with its coordinates permuted.

The canonical change is -0.017177, standard error 0.001953 and interval [-0.021008, -0.013346]. The reflected change is -0.055917, standard error 0.002119 and interval [-0.060072, -0.051762]. The equal-orientation change is -0.036547, with standard error 0.000774 and interval [-0.038064, -0.035029]. These estimates concern normalized block-count squared errors, not recognition percentages or a quality-rating scale (Stevens, 1946).

Because both orientation-specific sample means improve here, any fixed convex weighting of those two sample-average losses also improves. This broader average comparison follows from the observed negative changes at both endpoints, not from reflection consistency alone. It need not hold for another dataset, fitted predictor or loss. At the image level, an application-specific weighting can still favor the original rule, as the individual deterioration counts demonstrate.

Canonical error increases for 783 individual images despite its lower overall mean. Reflected error increases for 397. These counts use a numerical tolerance of 1e-12 for identifying a strict increase. They demonstrate why a favorable average does not imply imagewise dominance in the operational orientation. Influence-oriented analysis would examine which images contribute those increases before deciding whether the mean tradeoff is acceptable (Cook, 1977).

The largest equivariance residual in the stored calculation is zero, and the largest squared-loss identity residual is approximately 1.11e-16. Floating-point analysis explains the need to distinguish exact mathematical statements from their finite-precision checks (Higham, 2002; Goldberg, 1991). These residuals validate the implemented bookkeeping at the observed precision; they do not bound annotation error or establish semantic fidelity of the reconstructed digit.

Where this result sits among restoration methods

Modern inpainting methods operate far beyond four block-count targets. RePaint and LaMa address missing-image synthesis with different modeling choices (Lugmayr et al., 2022; Suvorov et al., 2021). Gated, partial and edge-connected inpainting explicitly organize how image information is used around missing content (Yu et al., 2018b; Liu et al., 2018; Nazeri et al., 2019). Our result provides no ranking among them because none was executed and their image outputs were not collected.

Architectural and generative tools also differ from the frozen affine map. U-Net describes an encoder-decoder approach to structured image prediction (Ronneberger et al., 2015), while diffusion formulations and implicit sampling concern learned generative processes (Ho et al., 2020; Song et al., 2020). An elementary reflection average may be definable for many predictors, but its effect on a stochastic sampler would require a declared coupling of random draws and an appropriate empirical evaluation.

Matrix-completion theory supplies other assumptions about recovering missing entries (Candes and Plan, 2010; Mazumder et al., 2010). Statistical imputation methods, including chained equations and nonparametric mixed-type imputation, address yet other data structures (Buuren and Groothuis-Oudshoorn, 2011; Stekhoven and Bühlmann, 2012). None of those theories turns our four-feature recovery into evidence that missing strokes or semantic details have been identified uniquely. The visible-to-hidden mapping remains a conditional approximation.

Measurement-error reasoning is useful because the scored block values and an intended historical inscription are different constructs (Carroll et al., 2006). Information-theoretic language likewise does not

make an averaging operation a source of new observations (Shannon, 1949). Both routes use rearrangements of the same sixty visible features. Any gain comes from constraining and combining the existing predictor's responses, not from acquiring extra evidence about the concealed block.

An appropriate next test depends on the deployment orientation

If a real restoration task encounters only canonical images, the canonical outcome should be primary. If orientation is genuinely uncertain and the two reflected states are equally plausible, the paired objective may be relevant. Our artificial fifty-fifty mixture is declared rather than inferred from an application. No orientation-frequency estimate is supplied by these digit files.

A future study could test a prespecified orientation distribution on independent collections, with writer-aware uncertainty and human or task-specific outcomes. Validation must follow that intended transfer boundary (Roberts et al., 2017). Such work should also distinguish a mathematical consistency constraint from an improvement learned from additional training examples, because they imply different information requirements and different sources of uncertainty.

The saved artifacts contain the operator identity, reflection permutation, original-image row, four loss columns and the two arithmetic checks. Documentation supports interpretation of those fields (Gebru et al., 2021), and reusable provenance links them to the official held partition (Wilkinson et al., 2016). For this frozen map, exact reflection consistency lowered both orientation-average errors, while many canonical images still deteriorated. The paired guarantee and the observed canonical tradeoff are consequently reported as separate findings.

References

1. Stekhoven, Daniel J.; Bühlmann, Peter. (2012). MissForest-non-parametric missing value imputation for mixed-type data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr597.
2. Ho, Jonathan; Jain, Ajay; Abbeel, Pieter. (2020). Denoising Diffusion Probabilistic Models. arXiv:2006.11239. https://arxiv.org/abs/2006.11239.
3. 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.
4. 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.
5. 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.
6. 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.
7. Candes, Emmanuel J; Plan, Yaniv. (2010). Matrix Completion With Noise. Proceedings of the IEEE. https://doi.org/10.1109/jproc.2009.2035722.
8. Higham, Nicholas J.. (2002). Accuracy and Stability of Numerical Algorithms: Second Edition. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718027.
9. 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.

10. 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.
11. Goldberg, David. (1991). What every computer scientist should know about floating-point arithmetic. ACM Computing Surveys. https://doi.org/10.1145/103162.103163.
12. Shannon, C.E.. (1949). Communication in the Presence of Noise. Proceedings of the IRE. https://doi.org/10.1109/jrproc.1949.232969.
13. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
14. 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.
15. Mazumder, Rahul; Hastie, Trevor; Tibshirani, Robert. (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices. Journal of Machine Learning Research 11, 2287-2322. https://jmlr.org/papers/v11/mazumder10a.html.
16. Song, Jiaming; Meng, Chenlin; Ermon, Stefano. (2020). Denoising Diffusion Implicit Models. arXiv:2010.02502. https://arxiv.org/abs/2010.02502.
17. Vincent, Pascal; Larochelle, Hugo; Bengio, Yoshua; Manzagol, Pierre-Antoine. (2008). Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th international conference on Machine learning - ICML '08. https://doi.org/10.1145/1390156.1390294.
18. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *Lecture Notes in Computer Science*, 2015, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.
19. Buuren, Stef van; Groothuis-Oudshoorn, Karin. (2011). mice: Multivariate Imputation by Chained Equations inR. Journal of Statistical Software. https://doi.org/10.18637/jss.v045.i03.
20. 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.
21. 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.
22. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.
23. Hoerl, Arthur E.; Kennard, Robert W.. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. https://doi.org/10.1080/00401706.1970.10488634.
24. Alpaydin, E.; Kaynak, C. (1998). Optical Recognition of Handwritten Digits [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C50P49.
25. Stevens, S. S.. (1946). On the Theory of Scales of Measurement. Science. https://doi.org/10.1126/science.103.2684.677.
26. 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.
27. 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.
28. 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.
29. Golub, G. H.; Reinsch, C. (1970). Singular value decomposition and least squares solutions. Numerische Mathematik, 14(5), 403-420. https://doi.org/10.1007/BF02163027.
30. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.