
Neural Phase Correlation

Cole Reynolds
Weyl Labs
cole@weyllabs.com

Abstract

Correspondence is fundamentally relational: it seeks the unknown transformation between two observations of a common scene, not the content of either. Yet the dominant learning-based methods do not represent the transformation as a first-class object in the architecture. They encode each image independently and let a learned similarity function or a deep decoder discover the mapping implicitly. Phase correlation is the canonical exception, measuring the inter-image relationship directly in the Fourier domain, but the rigidity of its fixed basis confines it to global translation.

We introduce a learned generalization of phase correlation that lifts this restriction by learning the basis on which the transformation decomposes. The same algebraic primitive extends to dense non-rigid deformations and to unitary dynamics. On the ACDC cardiac-MRI benchmark the framework matches or exceeds prior published baselines on both registration directions. On CAMUS echocardiography it matches state-of-the-art without auxiliary scoring or adaptive-smoothness mechanisms. Applied to time-evolved wavefunction pairs of the 1-D quantum harmonic oscillator, the same framework recovers the Hermite-function eigenstates and the quantized energy levels of the unknown Hamiltonian from observation pairs alone.

1 Introduction

Dense correspondence is a foundational task across imaging domains, establishing a per-pixel mapping between two views of related content. Methods for this problem fall broadly into two architectural traditions. Descriptor-based methods, from classical SIFT[1] to learned local descriptors[2], learned and dense feature matchers[3, 4], and cost-volume networks for optical flow[5–8], all encode each image independently and infer correspondence from a similarity computation between feature vectors. Concatenation-based registration networks[9, 10], dominant in medical imaging, take the channel-stacked image pair as input to a learned encoder and regress the displacement field directly, never constructing an explicit similarity. Phase correlation [11] is the structural exception: it measures the inter-image transformation directly in the frequency domain, never constructing a descriptor and never relying on a learned similarity. However, its applicability is bounded by the rigidity of the prescribed Fourier basis: each cross-power-spectrum peak encodes a single shift parameter, confining the method to global translation.

Phase correlation’s rigidity is well known; less remarked is what its rigidity buys. When phase correlation fails, the diagnosis is principled: the shift theorem assumed a transformation it cannot represent. This kind of structural transparency, in which failure modes are traceable to specific violated assumptions, has no analog in the descriptor- and concatenation-based architectures that have come to dominate dense correspondence. When the primitive of these methods fails, the response is typically additive, not corrective. While equivariant and geometric deep learning [12–15] have advanced the structural agenda in adjacent settings, dense correspondence has gone the other way: the architectural commitment to independent encoding has been amplified across two decades of work, yet the alternative of letting the architecture encode part of the transformation directly has remained largely unexplored.

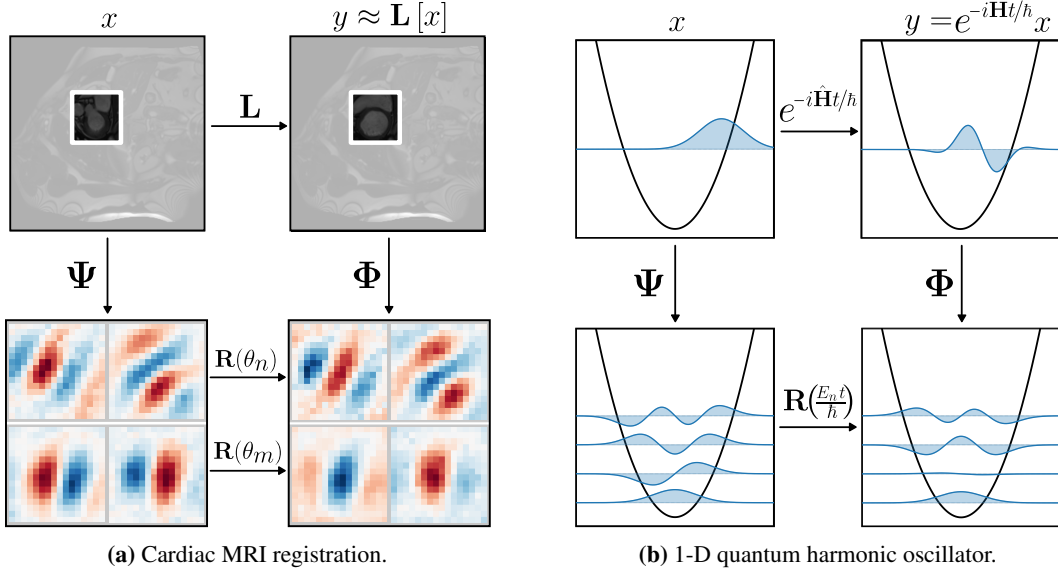


Figure 1: For filter-pair index n , the projection of the moving patch $\psi_n^T x$ and the projection of the fixed patch $\phi_n^T y$ lie in a 2-D subspace on which the local transformation acts as a planar rotation $\mathbf{R}(\theta_n)$ (Eq. (1), Eq. (4)). **Left:** cardiac MRI registration. **Right:** time-evolution of the 1-D quantum harmonic oscillator, with $\mathbf{L} = e^{-i\hat{H}t/\hbar}$ and $\theta_n = E_n t/\hbar$.

The natural generalization, then, is to keep the relational primitive but learn the basis on which it operates. We introduce a framework that recovers the per-subspace rotations of the operator relating two observations, replacing the fixed Fourier basis of classical phase correlation with two learned bases, Ψ and Φ (Figure 1). Each pair of filters learns to span a 2-D subspace on which the local operator acts as a planar rotation $\mathbf{R}(\theta)$, the direct generalization of the per-frequency translation phase that classical phase correlation extracts.

The framework’s central mechanism is a closed-form residual r_k^2 (Eq. (2)) that tests, per-location and per-subspace, whether the assumed rotation structure actually holds. The residual gates the architecture in two ways: at inference, it masks subspaces where the structure breaks down, preventing them from corrupting the displacement estimate; during training, top-K selection shapes the filters toward specialists for specific transformations rather than generalists for every transformation (Figure 3). Because the residual is derived from the algebraic model itself, it has no analog in descriptor-based pipelines, and its magnitude reveals where the structural assumption holds and where it fails.

The encoder is task-agnostic: filter pairs Ψ , Φ and per-pair rotation angles $\{\theta_n\}$ encode the local transformation regardless of what the transformation represents. The decoder that consumes this encoding is task-specific. For diffeomorphic image registration, the angles drive a dense displacement field $\varphi(\mathbf{r})$ such that $y(\mathbf{r}) \approx x(\mathbf{r} + \varphi(\mathbf{r}))$, turning the phase-correlation primitive into a fully differentiable, dense-registration module. For unitary time-evolution $e^{-i\hat{H}t/\hbar}$ of quantum states, the angle θ_n is the eigenenergy phase $E_n \tau/\hbar$ for some elapsed time τ , and the filters themselves reconstruct the wavefunction at τ in closed form (Section 4.3).

Contributions.

- A learned generalization of phase correlation that replaces the fixed Fourier basis with paired filter sets (Ψ , Φ) whose 2-D spans align with invariant subspaces of the local orthogonal operator, extending the phase-correlation primitive from global translation to dense, spatially-varying correspondence under non-rigid deformation.
- A closed-form, per-location residual r_k^2 (Eq. (2), with necessity and almost-sure sufficiency proved in Appendix B) that tests whether each filter-pair span is invariant under the local transformation. The residual masks non-invariant subspaces at inference, shapes the filter set toward consistent specialists during training via per-pair top-K selection (Appendix E, Figure 3), and serves as a built-in diagnostic revealing where the structural assumption holds and where it breaks down.

- A non-autonomous ODE on the displacement field that re-evaluates the model against the progressively warped moving image at each time step. Coupled to the residual mask, the active subspaces shift across integration time, recovering deformations that a single forward pass cannot.
- State-of-the-art on ACDC cardiac MRI (highest published Dice on the harder ED→ES direction; highest across-direction average under matched-loss training) and a match to the strongest CAMUS baseline without its auxiliary scoring or adaptive-smoothness machinery. Deformation breadth: validation on 2-D brain MRI slices and sub-meter synthetic-aperture-radar imagery (Appendix F), where it outperforms classical phase correlation.
- Applied to time-evolved wavefunction pairs of the 1-D quantum harmonic oscillator with randomized, unobserved time intervals τ , the same primitive recovers the Hermite-function eigenstates and the quantized energy levels of the unknown Hamiltonian from observation pairs alone (Section 4.3).

More broadly, this work generalizes phase correlation into a tool for recovering orthogonal operators from observation pairs.

2 Related Work

2.1 Medical image registration

Learned medical registration has converged on a similarity-driven paradigm: from DIRNet [16] through VoxelMorph [9] and its successors [10, 17–20], two images are stacked as input channels of a learned encoder that regresses a displacement field, inheriting the similarity-driven optimization of classical iterative methods.

2.2 Spectral and operator learning for dynamical systems

Outside image correspondence, several lines learn structural representations of dynamical operators from observation data, each committing to a particular mediating structure: a learned linear latent in the Koopman-net approaches [21, 22]; a variational wavefunction trained against a known Hamiltonian in neural-network quantum states [23]; a PDE residual at collocation points in physics-informed neural networks [24]; and a parameterized Hamiltonian ansatz [25]. Our framework instead consumes observation pairs directly, learning only the per-pair 2-D invariant-subspace decomposition from which the spectrum follows.

2.3 Learned Transformation Representations

Our encoder inherits its bilinear construction from a line of work originating with Memisevic and Hinton [26], whose three-way multiplicative connections represented inter-image transformations directly but at cubic parameter scaling. The three-way tensor was subsequently factored into per-image filter banks combined multiplicatively at the response level [27], introducing the parallel-bank construction we adopt. They demonstrated empirically that training on global shifts produces Fourier filter pairs in approximate quadrature, with the frequency-domain interpretation made explicit. The factored interaction was reinterpreted as detection of rotation angles within shared invariant eigen-subspaces of the transformation operator [28, 29]. This line demonstrated its principles for transformation detection and Eulerian-styled reconstruction on small image pairs. We extend the bilinear primitive to dense, spatially-varying correspondence at full resolution and add a closed-form per-subspace residual that tests invariance directly.

3 Method

For clarity, we describe the framework in image-registration language throughout this section and the registration experiments; Section 4.3 re-applies the same algebraic structure with the operator no longer an image warp, with notation adapted to wavefunctions and time evolution. Readers unfamiliar with classical phase correlation may find the brief overview in Appendix A useful before reading this section.

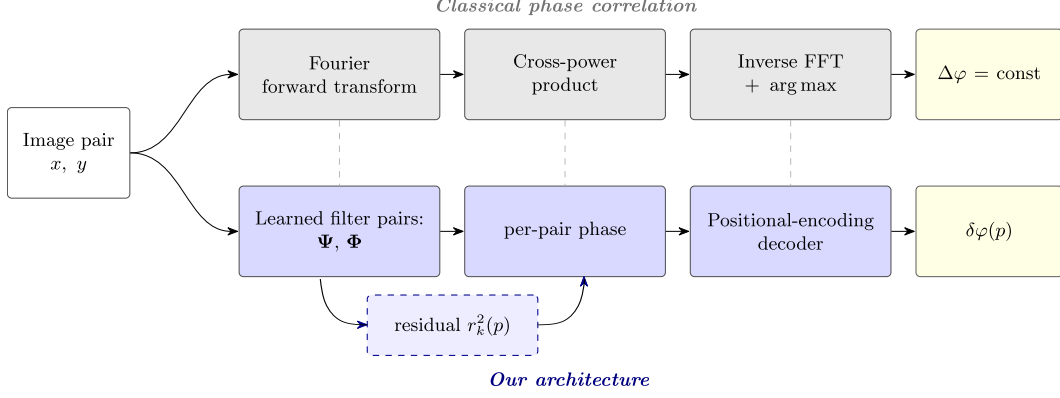


Figure 2: The proposed architecture generalizes the three stages of classical phase correlation. The per-pair invariance residual $r_k^2(p)$ (dashed) is a parallel branch with no classical analog: it is computed from the filter-pair responses and yields a top- K mask that selects which filter pairs contribute per-pair phase to the decoder.

Table 1: Stage-by-stage correspondence between classical phase correlation and the proposed architecture.

Stage	Classical phase correlation	This architecture
Forward transform	Fourier + conjugate (fixed)	Filter pairs ψ_k, ϕ_k (learned)
Per-frequency output	Re/Im of \hat{x}, \hat{y}	Pair responses $\psi_k^T x_k, \phi_k^T y_k$
Cross-power product	$\hat{x} \hat{y}$	Bilinear interaction of pair responses z_k
Phase extracted	Translation angle $\omega_k \cdot \Delta\varphi$	$(\cos \theta_k, \sin \theta_k, z_k)$
Transformation model	Global pure translation	Spatially-varying orthogonal L_p
Inverse/integration	Inverse FFT \Rightarrow 2-D plane with delta spike	PE residual branch (learned)

Image correspondence asks "where does this pixel go?". In the discrete limit it is a permutation P , and permutation matrices are orthogonal ($P^T P = I$); the patch-level model $L_p \in O(n)$ that follows is the natural continuous relaxation.

Let $x(p), y(p) \in \mathbb{R}^{2C}$ denote the vectorized sub-image patches of the moving and fixed images centered at location p . We model the patch transformation at each location as orthogonal, an approximation that holds locally for small displacements in the registration setting,

$$y(p) = L_p x(p), \quad L_p = \mathbf{U} \begin{bmatrix} R(\theta_1) & & \\ & \ddots & \\ & & R(\theta_C) \end{bmatrix} \mathbf{U}^T, \quad \mathbf{U} = [\hat{u}_1, \dots, \hat{u}_{2C}]. \quad (1)$$

Here \mathbf{U} is the (unknown, location-dependent) eigenbasis of L_p , and the angles θ_c likewise depend on p (notation suppressed for brevity); each $R(\theta_c) \in SO(2)$ is a planar rotation block on the 2-D subspace spanned by the c -th column pair of \mathbf{U} . The model interacts with the image pair only through the learned filter banks Ψ and Φ , which are themselves location-independent. The encoder learns column-pair subspaces ψ_k and ϕ_k whose spans align with the eigen-subspaces of L_p at as many locations as possible (full decomposition in Appendix C); the residual $r_k^2(p)$ tests whether this alignment holds and gates the architecture accordingly. Classical phase correlation corresponds to the special case where \mathbf{U} is the spatial Fourier basis: the column pairs are the (\cos, \sin) components of each frequency, and the bilinear interaction reduces to the cross-power spectrum.

Per-location, per-subspace residual. Let $\psi_k, \phi_k \in \mathbb{R}^{n \times 2}$ denote the k -th moving- and fixed-image filter pairs of Ψ and Φ , each a 2-column matrix in patch space. Because the pairs are learned and the bank is finite, the spans of ψ_k and ϕ_k cannot align with an invariant subspace of L_p at every location. Since orthogonal operators preserve inner products, and $\phi_k^T y \approx \phi_k^T L_p x$ under the orthogonal model, a mismatch between $\psi_k^T x$ and $\phi_k^T y$ is direct evidence of mis-alignment. The residual

$$r_k^2(p) = (\|\phi_k^T y(p)\| - \|\psi_k^T x(p)\|)^2, \quad \psi_k^T x(p) \in \mathbb{R}^2, \quad \phi_k^T y(p) \in \mathbb{R}^2, \quad (2)$$

formalizes this test. When ψ_k and ϕ_k span a common subspace, orthogonality of L_p guarantees $r_k^2(p) = 0$ if and only if that span is invariant under L_p (Lemma 1 and Theorem 1 in Appendix B,

with sufficiency holding almost surely under mild distributional assumptions on the moving image). The geometrically-normalized form $\bar{r}_k^2(p) = r_k^2(p) / (\|\phi_k^T y\| \|\psi_k^T x\|)$, used as the masking criterion throughout, inherits the same vanishing conditions.

Residual as a masking mechanism. A differentiable top- K selection [30] retains the K filter pairs with the smallest residuals at each location, zeroing the remaining $(C - K)$ before they propagate deeper into the network and changing the supervision incentive: the filters are allowed to be "greedy". A pair sharply aligned on some inputs and misaligned on others is rewarded for the former and dropped for the latter. As a result, the bank converges toward specialists rather than generalists, with per-pair residuals at active locations an order of magnitude below the unmasked baseline (Figure 3).

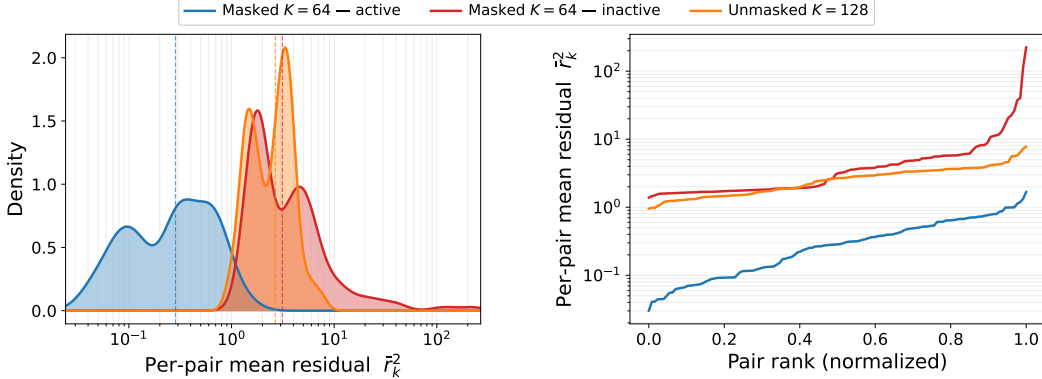


Figure 3: Per-pair mean residual \bar{r}_k^2 on held-out ACDC test pairs ($C = 128$; masked: $K = 64$, unmasked: $K = C$). **Left:** Histograms; **Right:** Sorted per-pair means. The masked distribution is split into *active* pairs (pair survived top- K) and *inactive* pairs (pair dropped). The histograms make the separation directly visible: at active pairs, masked residuals fall an order of magnitude below the unmasked baseline (median 0.25 vs 2.67); at inactive pairs, the same pairs would have matched or exceeded unmasked (median ~ 2.55). Full interpretation: Appendix E.

ODE instantiation. Each forward pass of the model M produces a displacement update $\delta\varphi$. Large deformations are recovered by iterating a non-autonomous ODE on the deformation field:

$$\frac{d\varphi}{dt} = M(x \circ \varphi(t), y), \quad \varphi(0) = \text{id}. \quad (3)$$

The model is re-evaluated at every step against the progressively-refined warped moving image $x \circ \varphi(t)$, distinct from stationary velocity-field formulations [31] that assume a fixed velocity over the integration interval. Re-evaluating at every step preserves the greediness of the per-pair top- K mask in time as well as in space: at each t the active filter pairs are determined by the current warped state, not the initial state. The active subspaces shift with integration time as a result, with low-frequency pairs typically dominating early steps to capture coarse displacement and higher-frequency pairs taking over for fine-scale refinement (Figure 7).

4 Experiments

4.1 Cardiac cine MRI (ACDC)

Under our preferred Normalized Cross Correlation (NCC) training recipe, our masked-residual architecture reports 0.863 average Dice on ACDC [32] and 0.843 on the harder ED \rightarrow ES direction, exceeding the strongest published baselines (DWCP at 0.837 / 0.813 under MSE). To isolate the architectural contribution from the loss change, we retrain VoxelMorph under our NCC recipe; our architecture exceeds matched-loss VoxelMorph + NCC, and VoxelMorph itself gains a comparable +0.029 from MSE to NCC, so the loss change does not asymmetrically favor our architecture (full loss-control breakdown in Appendix D.1). Under the exact Jian et al. [33] MSE protocol we exceed the published baselines on the harder ED \rightarrow ES direction (0.819; Table 3). Reported numbers are per-class mean $\pm 1\sigma$ standard deviation of per-pair test scores from seed 42; cross-seed stability across seeds 42, 43, 44 is reported separately in Appendix D.1.

Table 2: ACDC bidirectional registration, our 2-D protocol with NCC similarity (vs. MSE in Table 3); training recipe and metric definitions in Appendix D.1. The VoxelMorph + NCC row uses our identical recipe; only the architecture differs.

Method	ED→ES Dice	ES→ED Dice	Avg	SDlogJ	NDV /10k
Initial (no registration)	0.602 ± 0.148	0.602 ± 0.148	0.602	—	—
VoxelMorph + NCC	0.817 ± 0.103	0.867 ± 0.083	0.842	8.55 ± 1.35	83.1 ± 152.3
Ours (masked)	0.843 ± 0.104	0.883 ± 0.080	0.863	7.28 ± 1.79	76.4 ± 158.4

Table 3: ACDC bidirectional registration, Jian et al. [33] 2-D MSE protocol. Baselines as reported in their Table 5.

Method	ED→ES Dice	ES→ED Dice	Avg
Initial (no registration)	0.602 ± 0.148	0.602 ± 0.148	0.602
VoxelMorph [9]	0.788 ± 0.108	0.839 ± 0.083	0.814
TransMorph [10]	0.804 ± 0.097	0.855 ± 0.074	0.830
LKU-Net [18]	0.777 ± 0.107	0.822 ± 0.087	0.800
Mam-VXM [33]	0.800 ± 0.099	0.853 ± 0.076	0.827
Mam-TM [33]	0.804 ± 0.099	0.854 ± 0.077	0.829
Dual [33]	0.781 ± 0.110	0.831 ± 0.088	0.806
DWP [33]	0.803 ± 0.109	0.854 ± 0.085	0.829
DWCP [33]	0.813 ± 0.103	0.861 ± 0.079	0.837
DWCPi [33]	0.812 ± 0.104	0.858 ± 0.084	0.835
Ours (masked)	0.819 ± 0.112	0.836 ± 0.103	0.828

The pairing has structural basis: NCC and our encoder share the same intensity-scale invariance, while MSE is sensitive to intensity bias and scaling that our normalization discards. Each method is reported at its structurally aligned loss (DWCP at 0.813 under MSE, ours at 0.843 under NCC), and the +0.030 Dice gap reflects the architectural payoff rather than a loss-pairing artifact. Representative registrations and residual diagnostics appear in Figures 4 and 5.

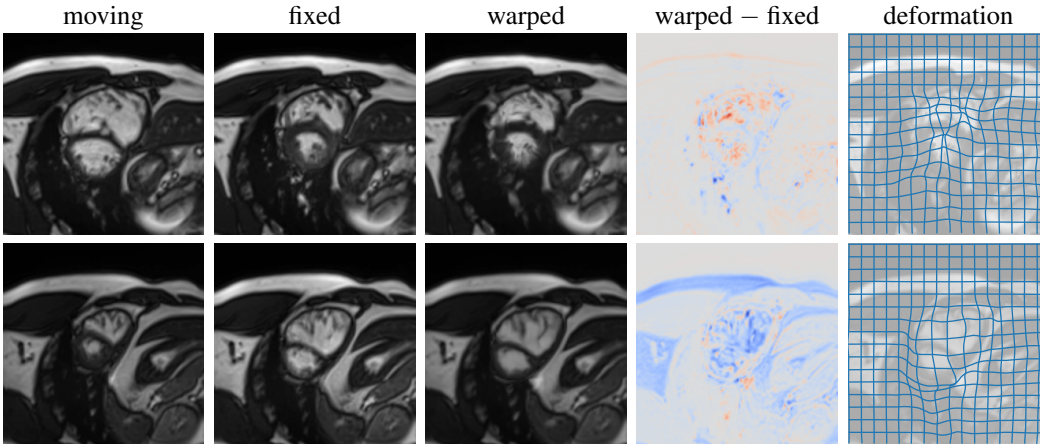


Figure 4: Registration examples on ACDC test pairs. Columns: moving image; fixed image; warped moving image; signed difference (warped – fixed) on a coolwarm scale; forward-displacement deformation field rendered as a deformed identity grid (the grid moves with the apparent image motion).

4.2 Echocardiography (CAMUS)

We evaluate on CAMUS [34] ED→ES reporting LV myocardium Dice as in Zhang et al. [17] (Appendix D.2). On CAMUS we deploy two cascaded stages of the same architecture, alternating within a single ODE integration: a coarse stage with effective kernel $k_f = 64$ at stride 8 for the first half of steps, then a fine stage with kernel $k_f = 32$ at stride 4 refining against the coarse-warped image.

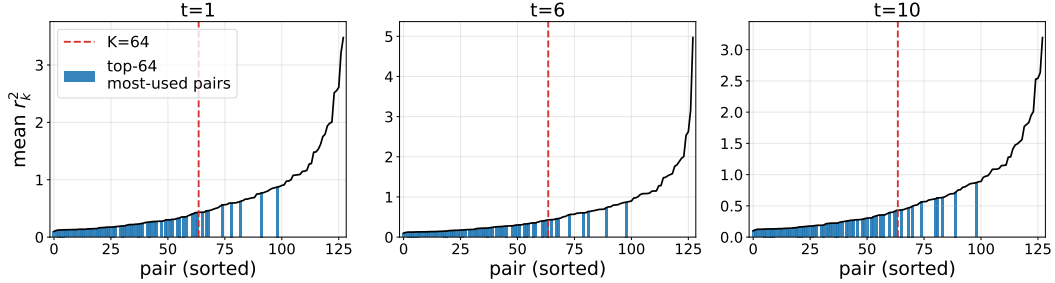


Figure 5: Per-pair sorted residuals at three ODE integration steps (first, middle, last) for a held-out ACDC test pair. Black curve: per-pair mean \bar{r}_k^2 sorted ascending. Blue fill: the $K=64$ pairs most frequently selected by the top- K mask at this step. Red dashed line: K cutoff. Each panel is independently y-scaled.

Table 4: CAMUS ED \rightarrow ES registration, LV myocardium Dice. Baselines as reported in Zhang et al. [17] Tables 1 and 5; ours is reported at seed 42, with cross-seed stability in Appendix D.2.

Method	Myo. Dice
Initial (no registration)	0.668
Elastix (classical)	0.802
DiffuseMorph + MSE	0.752 ± 0.087
TransMorph + MSE [10]	0.792 ± 0.061
VoxelMorph + MSE [9]	0.815 ± 0.056
VoxelMorph + AdaCS [17]	0.817 ± 0.054
Ours (masked, NCC, 2 stages)	0.820 ± 0.060

Our two-stage architecture matches VoxelMorph + AdaCS at 0.820 Dice vs 0.817 (within standard deviations), without the auxiliary correspondence-scoring estimator or adaptive-smoothness machinery; the residual mechanism described in Section 3 provides equivalent per-location reliability gating internally. Per-step residual diagnostics on a representative pair appear in Figure 7.

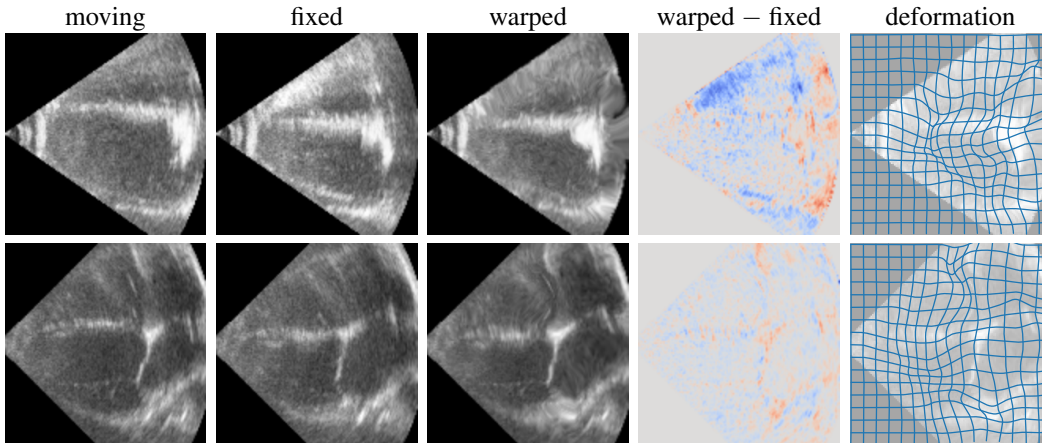


Figure 6: Registration examples on CAMUS test pairs. Columns: moving image; fixed image; warped moving image; signed difference (warped – fixed) on a coolwarm scale; forward-displacement deformation field rendered as a deformed identity grid (the grid moves with the apparent image motion).

4.3 Quantum Harmonic Oscillator: Eigenstate recovery from time pairs

The image-correspondence experiments above test the framework within a single problem family. To probe the algebraic generality of the primitive (filter pairs spanning 2-D invariant subspaces of an orthogonal operator), we apply the same architecture, with the spatial machinery stripped, to a problem in which the operator is no longer an image warp.

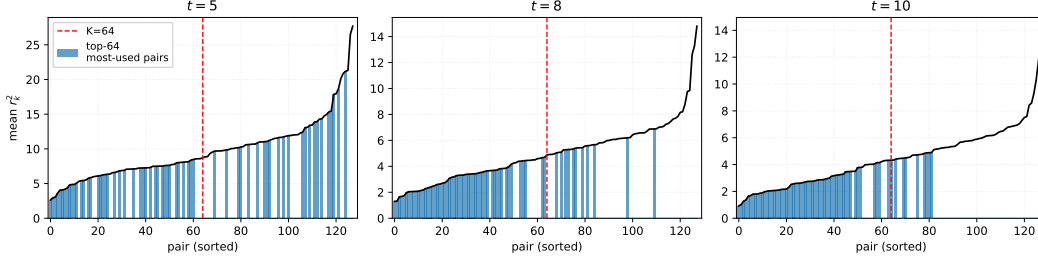


Figure 7: Per-pair sorted residuals at the first, middle, and last ODE integration steps of the fine stage for a held out CAMUS test pair (two-stage model, NCC; $C=128$ filter pairs). Black curve: per-pair mean \bar{r}_k^2 sorted ascending. Blue fill: the $K=64$ pairs most frequently selected by the top- K mask at this step. Red dashed line: K cutoff. Each panel is independently y-scaled.

For a time-independent Hamiltonian H with discrete spectrum (eigenvalue/vector pairs) $\{E_n, \psi_n\}$, the time-evolution operator $U(t) = \exp(-iHt/\hbar)$ is unitary and acts on each non-degenerate energy eigenspace as a planar rotation $R(E_n t/\hbar) \in SO(2)$, the same block-diagonal structure of Eq. (1) with $L = U(t)$. The filter-pairs learn the spatial representation of the eigenstates, and the bilinear interaction extracts the per-pair phase $E_n \tau/\hbar$.

We pick the 1-D quantum harmonic oscillator (QHO) for the cleanest possible test: eigenstates and energies are known analytically (Hermite functions, $E_n = \hbar\omega(n + \frac{1}{2})$). Training pairs are random unit-norm complex superpositions $\Psi(0) = \sum_{n=0}^{15} c_n \psi_n^{\text{true}}$ time-evolved analytically to $\Psi(\tau)$, with each c_n sampled from a complex normal distribution and τ drawn uniformly per pair from $[0.05, \tau_{\text{max}}]$, $\tau_{\text{max}} \in \{0.6, 1.7\}$. The model never sees H , the eigenstates, the coefficients, or τ (only the wavefunction pair). The architecture consists of $K = 16$ learned 1-D filters ψ_k , a per-filter learned rotation rate E_k , and a small permutation-equivariant readout MLP that recovers τ ; the wavefunction at τ is reconstructed in closed form (Appendix D.3).

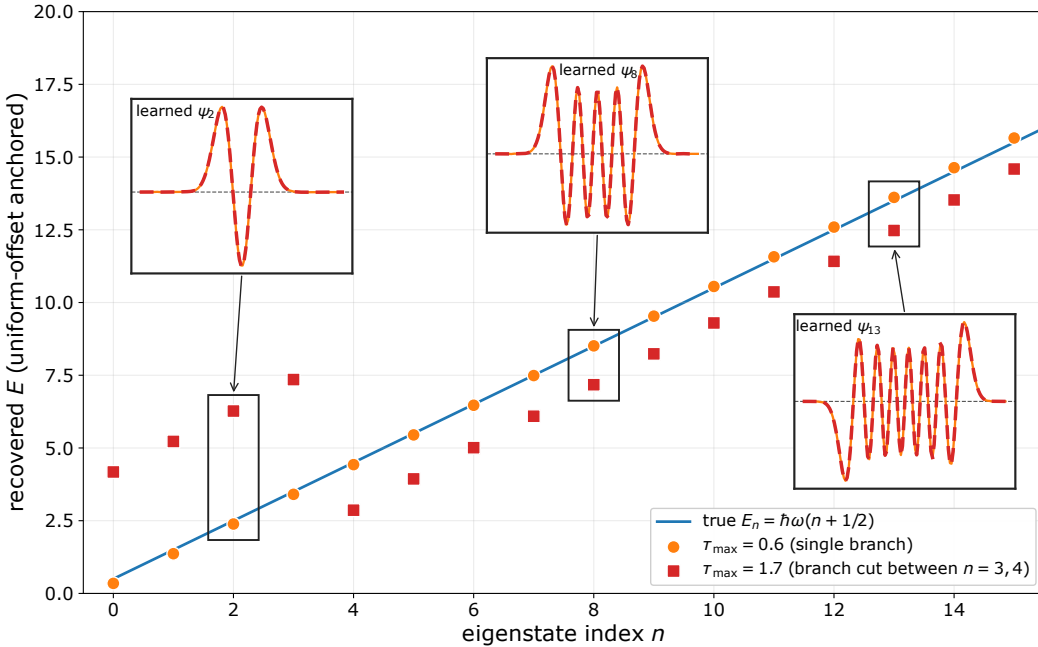


Figure 8: Cross-output asymmetry at $\tau_{\text{max}} = 1.7$. Recovered energy levels show fragmentation into two branches between the $n = 3$ and $n = 4$ states' boundary where $E_n \tau_{\text{max}}$ crosses 2π . Insets show learned eigenstates at representative n ; all learned eigenstates coincide with the analytic Hermite eigenstates with overlap 1.000. Full E_n and $\psi_n(x)$ recovery in Appendix D.3.

Figure 8 summarizes the results. Learned filters match the analytic Hermite eigenstates pointwise at both τ_{\max} regimes (median overlap 1.000). At $\tau_{\max} = 0.6$, recovered energies match the analytic values up to a single global offset; after calibrating against E_0 (read from the single-frequency oscillation of the matrix element $\langle \psi_0 | U(\tau) | \psi_0 \rangle$ (equivalently $\psi_0^T U(\tau) \psi_0$) across pair times, Appendix D.3; here $E_0 = 0.5$), per-filter absolute energy error falls below 0.05 for all n , and the held-out time stamps are recovered with Pearson correlation 1.000. From the recovered eigenstates and energies, the truncated Hamiltonian is reconstructed via the spectral decomposition $\hat{H} = \sum_n E_n^{\text{learned}} |\psi_n^{\text{learned}}\rangle \langle \psi_n^{\text{learned}}|$, with no parameterization of H or assumed functional form for the operator.

The framework’s two outputs, eigenstate recovery and eigenvalue recovery, exhibit different sensitivities to the time-interval distribution. Eigenstate recovery is τ -blind: learned filters continue to match the analytic eigenstates pointwise at both $\tau_{\max} = 0.6$ and $\tau_{\max} = 1.7$, with $\tau_{\max} = 1.7$ in the multi-wrap regime where $E_n \tau_{\max} > 2\pi$ for most filters. The structural primitive that identifies invariant subspaces (the residual) does not depend on τ , so the eigenvectors are unaffected by how aggressively the time interval is sampled. Eigenvalue recovery has a sharper, per-filter boundary at $E_n \tau_{\max} < 2\pi$ (Figure 8): at $\tau_{\max} = 0.6$ the ladder remains clean; at $\tau_{\max} = 1.7$ the spectrum fragments into two branches at the $n = 3 / n = 4$ boundary where $E_n \tau_{\max}$ crosses 2π . The mechanism is gauge-mediated: the reconstruction objective, NCC, is invariant to per-pair global phase, so filters with different wrap counts can settle on different unwrap branches without penalty. Eigenstate insets in Figure 8 show that the same training run that fragments the spectrum recovers ψ_2 (below the branch cut), ψ_8 and ψ_{13} (both above) pointwise.

5 Conclusion

We introduced a learned generalization of phase correlation that extends the primitive from the global-translation regime in which it has always operated [11] to dense, spatially-varying non-rigid deformation, and demonstrated its algebraic generality on unitary time-evolution where the same architecture recovers Hamiltonian eigenstates and energies from observation pairs alone. The per-pair residual and filter-bank visualization expose quantities interpretable within the model’s algebraic structure. The same algebraic ingredient that solves cardiac registration recovers a quantum Hamiltonian.

The structural commitment fixes the form of the answer in each setting. What the network learns is what the structure cannot determine analytically: which subspaces are locally invariant, how to assemble per-subspace rotations into a coherent output, and how to absorb the approximation gap when the operator is only locally orthogonal. The descriptor-based paradigm asks the network to discover correspondence from scratch; here, the network is responsible only for the residual irregularities of a structure that already encodes most of the task. Limitations are discussed in Appendix G.

References

- [1] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [2] Jerome Revaud, Cesar De Souza, Martin Humenberger, and Philippe Weinzaepfel. R2D2: Reliable and repeatable detector and descriptor. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [3] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [4] Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. RoMa: Robust dense feature matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [5] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Häusser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [6] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

- [7] Zachary Teed and Jia Deng. RAFT: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision (ECCV)*, 2020.
- [8] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [9] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.
- [10] Junyu Chen, Yong Du, Yufan He, William P. Segars, Ye Li, and Eric C. Frey. TransMorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, 82:102615, 2022.
- [11] Charles D. Kuglin and Donald C. Hines. The phase correlation image alignment method. In *IEEE Conference on Cybernetics and Society*, pages 163–165, 1975.
- [12] Taco S. Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning (ICML)*, 2016.
- [13] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [14] Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [15] Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [16] Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis (DLMIA), MICCAI Workshops*, pages 204–212. Springer, 2017.
- [17] Xiaoran Zhang, John C. Stendahl, Lawrence H. Staib, Albert J. Sinusas, Alex Wong, and James S. Duncan. Adaptive correspondence scoring for unsupervised medical image registration. In *European Conference on Computer Vision (ECCV)*, 2024.
- [18] Xi Jia, Joseph Bartlett, Tianyang Zhang, Wenqi Lu, Zhaowen Qiu, and Jinming Duan. U-Net vs transformer: Is U-Net outdated in medical image registration? In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2022.
- [19] Tao Guo, Yinuo Wang, Shihao Shu, Weimin Yuan, Diansheng Chen, Zhouping Tang, Cai Meng, and Xiangzhi Bai. MambaMorph: A mamba-based framework for medical MR-CT deformable registration. *arXiv preprint arXiv:2401.13934*, 2024.
- [20] Xi Jia, Wenqi Lu, Xinxing Cheng, and Jinming Duan. Decoder-only image registration. *IEEE Transactions on Medical Imaging*, 2024.
- [21] Bethany Lusch, Nathan Kutz, J. and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.
- [22] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems, 2017.
- [23] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.
- [24] Henry Jin, Marios Mattheakis, and Pavlos Protopapas. Physics-informed neural networks for quantum eigenvalue problems, 2022.
- [25] Timothy Kao et al. Solving the quantum many-body Hamiltonian learning problem with neural differential equations, 2024.
- [26] Roland Memisevic and Geoffrey E. Hinton. Unsupervised learning of image transformations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [27] Roland Memisevic and Geoffrey E. Hinton. Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural Computation*, 22(6):1473–1492, 2010.

- [28] Roland Memisevic. On multi-view feature learning. In *International Conference on Machine Learning (ICML)*, 2012.
- [29] Roland Memisevic. Learning to relate images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1829–1846, 2013.
- [30] Yujia Xie, Hanjun Dai, Minshuo Chen, Bo Dai, Tuo Zhao, Hongyuan Zha, Wei Wei, and Tomas Pfister. Differentiable top-k with optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 20520–20531, 2020.
- [31] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-Euclidean framework for statistics on diffeomorphisms. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 924–931. Springer, 2006.
- [32] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, 2018.
- [33] Bailiang Jian, Jiazhen Pan, Rohan Jena, Morteza Ghahremani, Huaqi Bell Li, Daniel Rueckert, Christian Wachinger, and Benedikt Wiestler. Disentangling progress in medical image registration: Beyond trend-driven architectures towards domain-specific strategies. *Medical Image Analysis*, 2025.
- [34] Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Østvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan D’hooge, Lasse Løvstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Transactions on Medical Imaging*, 38(9):2198–2210, 2019.
- [35] I. M. Gel’fand. *Lectures on Linear Algebra*. Dover Publications, New York, 1989. Translated from the Russian by A. Shenitzer.
- [36] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.
- [37] Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007.
- [38] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.
- [39] Lintao Li, Linyu Han, H. Y. Ye, Yiting Xiang, and Tian Zhang. Deep learning in remote sensing image matching: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 225:88–112, 2025.

A Classical phase correlation

Phase correlation [11] is a Fourier-domain technique for estimating the global translation between two signals. We illustrate the algorithm in the continuous 1-D setting for notational simplicity.

Setup. Let $f, g : \mathbb{R} \rightarrow \mathbb{R}$ be two functions related by a global translation $c \in \mathbb{R}$, i.e. $g(x) = f(x - c)$. The goal is to recover c from the pair (f, g) alone.

Fourier shift theorem. With the convention $\hat{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$ and $f(x) = \int_{-\infty}^{\infty} \hat{f}(k) e^{ikx} dk$ for the Fourier and inverse Fourier transforms of f , respectively, we have

$$\begin{aligned}
 \hat{g}(k) &= \int_{-\infty}^{\infty} g(x) e^{-ikx} dx \\
 &= \int_{-\infty}^{\infty} f(x - c) e^{-ikx} dx \\
 &= \int_{-\infty}^{\infty} f(u) e^{-ik(u+c)} du \quad \text{for } u = x - c \\
 &= e^{-ikc} \int_{-\infty}^{\infty} f(u) e^{-iku} du \\
 &= \hat{f}(k) e^{-ikc}.
 \end{aligned}$$

The translation is encoded entirely in the spatial frequency phase of \hat{g} relative to \hat{f} .

Cross-power spectrum and phase normalization. Multiplying $\hat{g}(k)$ by the complex conjugate $\overline{\hat{f}(k)}$ yields

$$\overline{\hat{f}(k)} \hat{g}(k) = |\hat{f}(k)|^2 e^{-ikc}.$$

The amplitude factor $|\hat{f}(k)|^2$ depends on the spectral content of f ; only the phase e^{-ikc} carries the translation. Normalizing by magnitude removes the amplitude and leaves a pure phase,

$$A(k) = \frac{\overline{\hat{f}(k)} \hat{g}(k)}{|\hat{f}(k) \hat{g}(k)|} = e^{-ikc}.$$

Peak localization. The inverse Fourier transform of A is a Dirac delta at the displacement,

$$\int_{-\infty}^{\infty} e^{-ikc} e^{ikx} dk = \int_{-\infty}^{\infty} e^{ik(x-c)} dk = \delta(x-c).$$

Thus applying $\arg \max$ to $\delta(x-c)$ or integrating $\int_{-\infty}^{\infty} x \delta(x-c) dx$ recovers c exactly when the global-translation assumption holds.

Vectorized discrete formulation. The same construction in matrix form, with $\mathbf{f}, \mathbf{g} \in \mathbb{R}^N$ the discrete samples of f and g on a grid of N points. The Fourier transform becomes the $N \times N$ DFT matrix \mathbf{F} with entries $\mathbf{F}_{nm} = e^{-i2\pi nm/N}/\sqrt{N}$; on real inputs its output satisfies the conjugate-symmetry relation $(\mathbf{F}\mathbf{f})_{N-n} = \overline{(\mathbf{F}\mathbf{f})_n}$. The translation $g(x) = f(x-c)$ becomes

$$\mathbf{g} = \mathbf{T}_c \mathbf{f}, \quad \mathbf{T}_c = \mathbf{F}^{-1} \mathbf{D}_c \mathbf{F},$$

where \mathbf{D}_c is diagonal with $\mathbf{D}_{c,nn} = e^{-ik_n c}$ and $k_n = 2\pi n/N$. The factorization is the matrix form of the *FFT* \rightarrow *phase shift* \rightarrow *IFFT* algorithm.

Translation is orthogonal. On real inputs \mathbf{T}_c is real-orthogonal: $\mathbf{T}_c^T \mathbf{T}_c = \mathbf{I}$. Orthogonal matrices are the real-valued counterpart of unitary matrices, with each complex-conjugate eigenvalue pair acting as a real 2×2 rotation block on the corresponding 2-D real subspace [35]; here, the conjugate pair $(e^{-ik_n c}, e^{+ik_n c})$ on the diagonal of \mathbf{D}_c becomes $R(k_n c) \in SO(2)$ on the cosine/sine pair at frequency k_n . \mathbf{T}_c thus has the same block-diagonal-with-rotations structure as the patch operator \mathbf{L}_p in Eq. (1), and is the formal counterpart of the ‘‘permutation matrices are orthogonal’’ observation in Section 3.

Spatially-varying displacement. When c is constant, \mathbf{T}_c is block-diagonal in the spatial harmonic basis: a single $R(k_n c)$ at every frequency. When c varies with x , $g(x) = f(x - c(x))$, the substitution that pulled c out of the integral fails: $u = x - c(x)$ no longer has a constant Jacobian, and

$$\hat{g}(k) = \int_{-\infty}^{\infty} f(x - c(x)) e^{-ikx} dx \neq \hat{f}(k) e^{-ikc}.$$

The matrix counterpart is that $\mathbf{T}_{c(\cdot)}$ is still real-orthogonal, but is no longer block-diagonal in the spatial harmonic basis: the action couples frequencies.

Learning the right basis. Spatially-varying displacement remains a real-orthogonal operation at each location — still block-diagonalizable into 2×2 rotations, just not by the spatial harmonic basis. Neural Phase Correlation seeks the basis \mathbf{U} in which the local transformation \mathbf{L}_p is block-diagonal with 2×2 rotation blocks (Eq. (1)). Classical phase correlation is the special case in which \mathbf{U} is the spatial harmonic basis, optimal only when c is constant. For spatially-varying $c(x)$, the diagonalizing basis depends on c itself; the learned filter pairs Ψ, Φ are the vehicle for recovering it from data.

B Residual

This appendix formalizes the basis for interpreting the per-location, per-subspace residual $r_k^2(p)$ of Section 3 as an inference-time validity signal for $\text{span}(\psi_k)$ under the local orthogonal operator L_p . We first analyze the residual in the idealized scenario in which both filter pairs span a common invariant subspace of L_p , establishing its invariance and swap symmetry. The realistic case in which alignment may fail is then characterized by Lemma 1 and Theorem 1.

Idealized scenario: two filter pairs, three rotations. The main-text residual is $r_k^2 = (\|\phi_k^T y\| - \|\psi_k^T x\|)^2$, with separate filter pairs for the moving image (ψ_k) and the fixed image (ϕ_k). Suppose both pairs span the k -th 2-D eigenspace of L_p with orthonormal columns:

$$\psi_k = U_k R(\theta_\psi), \quad \phi_k = U_k R(\theta_\phi),$$

where $U_k = [u_{2k-1}, u_{2k}]$ is the k -th 2-D eigenspace of L_p , with $L_p|_{U_k} = R(\theta_k) \in SO(2)$, and $R(\theta_\psi), R(\theta_\phi) \in SO(2)$ are the in-plane orientations of the two filter pairs relative to that eigenframe.¹ Three rotations are then in play, all in $SO(2)$:

- $R(\theta_\psi)$ and $R(\theta_\phi)$: the orientations of the moving and fixed filter pairs in the eigenframe U_k .
- $R(\theta_\psi - \theta_\phi)$: the inter-filter rotation between the moving and fixed pairs, parameterized by a single angle difference.

Norm equality and swap symmetry. Let $p_x = U_k^T x$ be the U_k -projection of x in eigenframe coordinates, and write $y = L_p x$. Each of the four projection norms reduces to $\|p_x\|$:

$$\begin{aligned} \|\psi_k^T x\| &= \|R(-\theta_\psi) p_x\| = \|p_x\|, & \|\phi_k^T y\| &= \|R(-\theta_\phi) R(\theta_k) p_x\| = \|p_x\|, \\ \|\psi_k^T y\| &= \|R(-\theta_\psi) R(\theta_k) p_x\| = \|p_x\|, & \|\phi_k^T x\| &= \|R(-\theta_\phi) p_x\| = \|p_x\|, \end{aligned}$$

since each is a planar rotation applied to p_x and $SO(2)$ preserves the Euclidean norm. Hence $r_k^2 = 0$, and the swap $(\psi_k, \phi_k) \rightarrow (\phi_k, \psi_k)$ gives $(\|\psi_k^T y\| - \|\phi_k^T x\|)^2 = 0$ by the same calculation: the residual is invariant under exchange of moving and fixed filter pairs in the ideal case.

Setup. Let $L \in O(n)$ with $n = k_f^2$, so $L^T L = I$, and assume each filter pair is orthonormal,

$$\psi_k^T \psi_k = \phi_k^T \phi_k = I_2,$$

and that the two pairs share a common span: $\text{span}(\psi_k) = \text{span}(\phi_k)$, equivalently $\phi_k = \psi_k R$ for some $R \in O(2)$ (this is the idealized structure; its failure is precisely what the residual measures). The two projectors then coincide,

$$P = \psi_k \psi_k^T = \phi_k \phi_k^T \in \mathbb{R}^{n \times n},$$

since $\phi_k \phi_k^T = \psi_k R R^T \psi_k^T = \psi_k \psi_k^T$. For a patch $x \in \mathbb{R}^n$ and its transformed counterpart $y = Lx$, the moving projection is $\psi_k^T x \in \mathbb{R}^2$, the fixed projection is $\phi_k^T y \in \mathbb{R}^2$, and

$$r^2 = (\|\phi_k^T y\| - \|\psi_k^T x\|)^2.$$

Invariance condition. $\text{span}(\psi_k)$ is invariant under L , i.e. $L \text{span}(\psi_k) \subseteq \text{span}(\psi_k)$, if and only if the projector P commutes with L :

$$[P, L] := PL - LP = 0.$$

Under this condition the restriction $L|_{\text{span}(\psi_k)}$ is itself orthogonal: for any $u, v \in \text{span}(\psi_k)$, $\langle Lu, Lv \rangle = \langle u, v \rangle$ since $L \in O(n)$ preserves inner products globally, so its restriction inherits the property. On the 2-D subspace $\text{span}(\psi_k)$, L therefore acts as a planar rotation $R(\theta)$ for some θ (the $SO(2)$ case) or a rotation composed with a reflection (the $O(2) \setminus SO(2)$ case); both preserve the Euclidean norm.

Lemma 1 (Necessity). If $\text{span}(\psi_k)$ is invariant under L (equivalently $[P, L] = 0$), then $r^2(x) = 0$ for every $x \in \mathbb{R}^n$.

Proof. Since L is orthogonal, $L^T = L^{-1}$. Expanding using $P = \phi_k \phi_k^T = \psi_k \psi_k^T$,

$$\|\phi_k^T y\|^2 = y^T P y = x^T L^T P L x = x^T L^{-1} P L x.$$

The commutativity $PL = LP$ gives $L^{-1} P L = P$, so

$$\|\phi_k^T y\|^2 = x^T P x = \|\psi_k^T x\|^2.$$

Taking square roots, $\|\phi_k^T y\| = \|\psi_k^T x\|$ and hence $r^2 = 0$. \square

Theorem 1 (Almost-sure sufficiency). Suppose $\text{span}(\psi_k)$ is not invariant under L , i.e., $[P, L] \neq 0$. If x is drawn from an absolutely continuous distribution on \mathbb{R}^n independently of L , then

$$\mathbb{P}(r^2(x) = 0) = 0.$$

¹Restricting to $SO(2)$ rather than $O(2)$ excludes a reflection of each filter frame; the norm-equality argument below goes through identically in the reflected case, so the residual's vanishing condition is unaffected.

Proof. The residual vanishes iff $\|\phi_k^T y\|^2 = \|\psi_k^T x\|^2$, i.e.,

$$x^T A x = 0, \quad A := L^{-1} P L - P.$$

A is symmetric. Using $P^T = P$, $L^T = L^{-1}$, and $(L^{-1})^T = L$,

$$A^T = L^T P^T (L^{-1})^T - P^T = L^{-1} P L - P = A.$$

$A \neq 0$. Left-multiplying A by L yields

$$L A = P L - L P = [P, L].$$

Since L is invertible, $A = 0 \iff [P, L] = 0$. By hypothesis $[P, L] \neq 0$, hence $A \neq 0$.

Thus $x \mapsto x^T A x$ is a nonzero quadratic form on \mathbb{R}^n . The set where it vanishes, $\{x : x^T A x = 0\}$, is the zero set of a nonzero polynomial in n real variables, and the zero set of any nonzero polynomial on \mathbb{R}^n has Lebesgue measure zero. Any continuous probability distribution on \mathbb{R}^n assigns probability zero to a measure-zero set, hence $\mathbb{P}(x^T A x = 0) = 0$. \square

Structure of the failure set. The vanishing set above can be described more concretely. Both P and $L^{-1} P L$ are rank-2 projections, so their difference A has rank at most four: the value of $x^T A x$ depends only on the components of x along at most four particular directions in \mathbb{R}^n , and is identically zero along the other $n - 4$.

C Architecture and implementation details

This appendix gives the full description of each stage of the architecture summarized in Section 3 and Figure 2.

C.1 Stage 1: Learned filter pairs

The orthogonal-operator model $y(p) = L_p x(p)$ (Eq. (1)) defines the setting: L_p admits a canonical decomposition into 2×2 rotation blocks that organize our generalization of the per-frequency structure of phase correlation. Because the invariant subspaces of a general orthogonal L_p depend on L_p itself, the fixed Fourier basis used by classical phase correlation is replaced with two banks of C learnable filter pairs, ψ_k and ϕ_k for $k = 1, \dots, C$, applied to the moving and fixed images via strided convolution (each in $\mathbb{R}^{n \times 2}$). The sole architectural constraint is unit ℓ_2 norm on each filter; orthonormality within a pair and alignment of its span with an invariant subspace of L_p are not imposed architecturally; they emerge from training against the task loss. Figure 1 shows representative learned filter pairs on cardiac MRI.

Under the ideal structure ($\text{span}(\psi_k)$ invariant under L_p , pair orthonormal), the restriction $L_p|_{\text{span}(\psi_k)}$ is a planar rotation $R(\theta_k(p)) \in SO(2)$. Each filter pair plays the role of a single classical frequency: it identifies a 2-D subspace on which the local transformation acts as a rotation by an angle $\theta_k(p)$. Memisevic and Hinton [27] established empirically that filter pairs in the factored bilinear architecture converge to Fourier harmonic pairs when trained on pure-translation data, so the Fourier-filter limit case in our reduction is realized by training rather than imposed.

C.2 Stage 2: Bilinear interaction on filter responses

Classical phase correlation forms the cross-power spectrum $\hat{y}(k) \cdot \overline{\hat{x}(k)}$ by multiplying per-frequency Fourier coefficients. Under our modeling assumption, the analogous operation is a *bilinear interaction on the filter-response subspace*, evaluated independently at each coarse-grid location p . Let $x_p, y_p \in \mathbb{R}^n$ denote the moving and fixed patches centered at p . The per-subspace bilinear interaction is

$$z_k(p) = \begin{pmatrix} (\psi_k^T x_p)^T (\phi_k^T y_p) \\ (\psi_k^T x_p)^T J (\phi_k^T y_p) \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (4)$$

where $\psi_k^T x_p, \phi_k^T y_p \in \mathbb{R}^2$ are the projections of the moving and fixed patches onto their respective filter pairs. Each component of $z_k(p)$ is a bilinear form in $(\psi_k^T x_p, \phi_k^T y_p)$: the symmetric inner product, and the skew-symmetric inner product induced by the planar 90-degree rotation J . The pair-product is computed entirely in the encoded filter-response space and never returns to image space; this is what we refer to as bilinear interaction.

Under $\mathbb{R}^2 \cong \mathbb{C}$ (encoding $v = (v_1, v_2) \in \mathbb{R}^2$ as $v_1 + iv_2 \in \mathbb{C}$), z_k is $\overline{\psi_k^T x_p} \cdot \phi_k^T y_p$, the real-vector form of the classical cross-power product, and its two components are proportional to $(\cos \theta_k, \sin \theta_k)$. The decoder receives a polar factorization: unit-circle direction (the per-subspace rotation angle) and magnitude separately $(\cos \theta_k, \sin \theta_k, |z_k|)$. Magnitude normalization inherits the classical phase-correlation cancellation of intensity bias; keeping magnitude as a separate channel lets downstream weighting attenuate a subspace’s contribution without corrupting its angle estimate.

C.3 Stage 3: Positional-encoding decoder

Classical phase correlation completes the pipeline by inverse-transforming the normalized cross-power spectrum to produce a 2-D plane with a Dirac peak at the displacement, then localizing that peak by $\arg \max$. We instead interpret the displacement vector as integrating the product of the nearly-Dirac peak and the coordinate grid. The decoder implements this interpretation through two intermediate spaces: a *gate space* \mathbb{R}^G in which per-pair features and positional encodings interact multiplicatively, and the *PE space* $\mathbb{R}^{n_{pe}}$ in which the displacement residual lives. In our released configurations the gate-space dimension equals the filter-pair count ($G = C$), though the two spaces are structurally distinct: \mathbb{R}^C is indexed by filter pair, while \mathbb{R}^G is the space in which pair features and positional encodings combine.

Let $\gamma \in \mathbb{R}^{n_{pe}}$ denote a NeRF-styled sinusoidal positional encoding of the coordinate grid [36]. The masked bilinear interaction at the current coarse-grid location yields a per-pair polar feature vector $z \in \mathbb{R}^{3C}$, stacking the triple $(\cos \theta_k, \sin \theta_k, |z_k|)$ over $k = 1, \dots, C$. A two-layer 1×1 MLP $\text{MLP}_z : \mathbb{R}^{3C} \rightarrow \mathbb{R}^G$ with hidden width B and a LeakyReLU between layers projects z into the gate space, while a single learned linear map $W_\gamma \in \mathbb{R}^{G \times n_{pe}}$ lifts γ into the same space; the two are multiplied element-wise (resembling the delta peak’s product with the coordinate grid in classical phase correlation), a second MLP maps the gated vector back to PE space, and a final readout produces the Cartesian displacement:

$$\begin{aligned} m &= \text{MLP}_z(z), & \gamma_g &= W_\gamma \gamma, \\ \delta\gamma &= \text{MLP}_{pe}(m \odot \gamma_g) - \gamma, & \delta\varphi &= \text{head}(\delta\gamma), \end{aligned} \tag{5}$$

where $m, \gamma_g \in \mathbb{R}^G$, $\delta\gamma \in \mathbb{R}^{n_{pe}}$, \odot is element-wise multiplication, and $\text{MLP}_{pe} : \mathbb{R}^G \rightarrow \mathbb{R}^{n_{pe}}$ is a two-layer 1×1 net with hidden width G and a LeakyReLU between layers (Table 5). The residual $\delta\gamma$ is the displacement representation in PE space; head maps it to a Cartesian displacement vector. The canonical readout is a single linear $\text{Conv}_{1 \times 1}(n_{pe}, 2)$; in this paper we use a depth-1 mini-UNet of the same input/output shape with 3×3 spatial kernel size, which couples neighboring spatial indices for additional regularization.

C.4 Training objective

All checkpoints are trained unsupervised against a similarity loss between the warped moving and fixed images, plus a smoothness regularizer on the displacement field and a Jacobian-determinant fold penalty for topology preservation. The full loss specification is reported per benchmark in Table 6. The ODE integration scheme used at inference is the one defined by Eq. (3).

C.5 Implementation details

The conceptual architecture is described in the preceding subsections of this appendix. Both released checkpoints (ACDC, CAMUS) instantiate the same UNet decoder architecture and differ only in a small number of capacity, regularization, and schedule choices appropriate to each modality. Across both benchmarks the encoder is a normalized convolution producing two banks of C filter pairs at coarse-grid stride s ; per-pair features and positional encodings interact in a gate space of dimension G before being mapped back to the PE space and read out as a Cartesian displacement. The depth-1 mini-UNet uses a hidden channel dimension of 16. Reported numbers are per-pair test-set mean $\pm 1\text{-}\sigma$ sample standard deviation from seed 42; cross-seed stability across seeds 42, 43, 44 is reported in the per-benchmark appendices.

Table 5: Architecture hyperparameters per benchmark.

	ACDC	CAMUS
Image size	128×128	128×128
Encoder kernel k_f	16	64 & 32
Encoder stride s	4	8 & 4
Coarse-grid resolution	32×32	32×32
Filter-pair bank C	128	128
Top- K mask	64	64
Gate-space dim G	128	128
Encoder MLP hidden B	128	128
Positional-encoding channels n_{pe}	256	256
Integration steps n_t	10	10

D Per-benchmark protocol details

This appendix gives data-, loss-, and evaluation-specific details for each of the three benchmarks. Architecture, optimizer, and per-benchmark hyperparameters are summarized in Tables 5–7; the prose below specifies what

Table 6: Loss and optimization hyperparameters per benchmark. Loss is $\mathcal{L} = \mathcal{L}_{\text{sim}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_J\mathcal{L}_{|J|} + \lambda_{\log J}\mathcal{L}_{\log J}$. All checkpoints use Adam.

	ACDC	CAMUS
\mathcal{L}_{sim}	NCC (window 17)	NCC (window 67)
Regularizer \mathcal{L}_{reg}	$\ \nabla\varphi\ ^2$	$\ \nabla\varphi\ ^2$
λ_{reg}	5×10^{-2}	2×10^{-1}
λ_J (Jac. det.)	100	1
$\lambda_{\log J}$	10^{-5}	10^{-2}
Learning rate	10^{-3}	10^{-3}
LR schedule	exp. $\gamma=0.997$	exp. $\gamma=0.996$
Batch size	50	50
Epochs	400	300
Weight decay	0	10^{-5}
Gradient clip (norm)	off	off
Augmentation	h/v flip w/ $p=0.5$	none
Pairs per epoch	full train set	full train set
Seeds reported	42, 43, 44	42, 43, 44

Table 7: Per-benchmark dataset protocol. Splits and pre-processing follow the original published protocols.

	ACDC	CAMUS
Source protocol	Jian et al. [33]	Zhang et al. [17]
Patient split (train/val/test)	80/20/50	300/100/100
Pairs	ES-ED, both directions	ES-ED, both 2CH & 4CH views
In-plane resampling	as in source	1.0 mm
Crop / canvas	128×128 ROI on seg. bbox	128×128 center-crop
Evaluation metric	Dice (RV / Myo / LV)	LV-myocardium Dice

those tables do not (split rules, evaluation metric definitions, cross-seed stability, comparison caveats). All datasets are publicly distributed for research use: ACDC [32] and CAMUS [34] via CREATIS (free registration), OASIS-1 [37] under the OASIS Data Use Agreement.

D.1 Cardiac MRI (ACDC)

Data and split (Jian et al. [33] §4.1.3). We use all 150 labeled ACDC subjects, split 80 train / 20 validation / 50 test at the patient level (seed 2023), yielding 506 / 190 / 194 annotated 2-D short-axis slices respectively ([33] reports 194 test pairs, but we obtained 394 which is more consistent with the split ratios and assume their test split is a subset of ours). We keep slices in which both ED and ES segmentations contain all three structures (RV, myocardium, LV), reproducing Jian et al. [33]’s reported train+val slice count of $506 + 190 = 696$. Each slice is cropped to a 128×128 region of interest centered on the union of the ED/ES segmentation bounding boxes and intensities are min-max normalized to $[0, 1]$ per slice.

Bidirectional training. Each iteration performs both forward (ED→ES) and backward (ES→ED) registrations through the same model, with the total loss averaged across directions, matching Jian et al. [33] §4.3.

Evaluation. Both directions are evaluated separately at test time. We report Dice over the three annotated structures (RV, myocardium, LV) per direction and as the across-direction average. $\text{SDlogJ} (\times 10^{-2})$ is the standard deviation of $\log(\det J + 3)$ following Jian et al. [33]’s implementation, lower indicating smoother volume changes. NDV (/10k) is the rate of pixels with non-positive Jacobian determinant per 10,000 foreground voxels (foreground = union of moving and fixed segmentations), lower indicating more diffeomorphic warps.

Loss-control breakdown. Under the NCC recipe above, our masked-residual architecture reaches 0.843 Dice on ED→ES, while VoxelMorph re-trained under the identical recipe reaches 0.817; the +0.026 gap is the architectural contribution. The same VoxelMorph re-training shows +0.029 NCC-over-MSE (0.788→0.817), and our own architecture shows a comparable +0.024 NCC-over-MSE (0.819→0.843). Both architectures gain similarly from NCC, so NCC does not asymmetrically favor our method.

Cross-seed stability (Table 2). The Table 2 numbers are reported at seed 42. Across three independent training seeds (42, 43, 44) under the recipe above, the cross-seed standard deviations are 0.001, 0.003, and 0.002

on Dice (ED→ES, ES→ED, and the across-direction average respectively), 0.16 on SDlogJ, and 2.9 on NDV. The reported deltas over DWCP [33] therefore exceed cross-seed variability by an order of magnitude in each direction.

D.2 Echocardiography (CAMUS)

Data and split. The CAMUS dataset [34] comprises 500 patients with paired apical two- and four-chamber views annotated at ED and ES with LV myocardium segmentation. We adopt the Zhang et al. [17] protocol: a random 300/100/100 patient-level split (seed 42) yielding 600/200/200 train/val/test 2-D image pairs after including both 2CH and 4CH views per patient. Images are resampled to 1.0 mm in-plane and center-cropped to 128×128 .

Evaluation. ED→ES Dice on LV myocardium, reported at seed 42, with cross-seed stability across seeds 42, 43, 44 characterised below. Zhang et al. [17] does not report standard deviations for the Initial and Elastix baselines.

Cross-seed stability (Table 4). The Table 4 number is reported at seed 42. Across three independent training seeds (42, 43, 44) under the recipe above, the LV-myocardium Dice values are 0.8204, 0.8190, and 0.8180 (cross-seed mean 0.8191, cross-seed standard deviation 0.001). The reported margin over VoxelMorph + AdaCS [17] is therefore within cross-seed variability, consistent with the parity claim in Section 4.2.

D.3 Quantum Harmonic Oscillator

This appendix supplements §4.3 with the full architectural and training specification for the QHO experiment, a discussion of failure modes at large τ , and positioning relative to existing literature.

Common setup. Position grid $x \in [-8, 8]$ with $N = 128$ points; $\hbar = \omega = m = 1$; the first $n_{\max} = 16$ Hermite-function eigenstates computed analytically via `scipy.special.hermite` and discretely L2-renormalized. Each training pair samples a random complex unit-norm coefficient vector $c \in \mathbb{C}^{16}$, forms $\Psi(0) = \sum_n c_n \psi_n^{\text{true}}$, and time-evolves analytically eigenstate-by-eigenstate to obtain $\Psi(\tau) = \sum_n c_n e^{-iE_n \tau / \hbar} \psi_n^{\text{true}}$.

Architecture and training. Sampler: τ_b drawn uniformly from $[\tau_{\min}, \tau_{\max}]$ per pair, not provided to the model; we report results at $\tau_{\max} \in \{0.6, 1.7\}$ (§4.3). Architecture: bank $\psi_k \in \mathbb{R}^N$ ($K = 16$, unit-normalized at use time); learned phase angles δ_k ; per-filter learned rotation rate E_k (K -vector); permutation-equivariant readout MLP \mathcal{M} that ingests $\{(E_k, \cos \theta_k(b), \sin \theta_k(b), |z_k(b)|)\}_{k=1}^K$ per pair, pools across the filter axis with mean+max, and outputs $\hat{\tau}_b \in [\tau_{\min}, \tau_{\max}]$ via sigmoid range-scaling (hidden width 32, $\approx 3.3k$ parameters). Reconstruction is closed-form in the diagonalized basis,

$$\hat{c}_k(b) = c_k(\Psi_0^{(b)}) \cdot \exp(-i E_k \hat{\tau}_b), \quad \hat{\Psi}(\tau_b) = \sum_k \hat{c}_k(b) \psi_k.$$

Total parameters: $KN + K + |\mathcal{M}| = 2,048 + 16 + 3,329 \approx 5.4k$. Loss: magnitude-NCC reconstruction against $\Psi(\tau_b)$ + residual (\mathcal{L}_{res} , Eq. (2)) + diversity (\mathcal{L}_{div} , off-diagonal Gram squared) + span coverage ($\mathcal{L}_{\text{span}} = \mathbb{E}[|\|\Psi\|^2 - \sum_k |c_k|^2|]$); all weights 1.0. Adam at 10^{-3} , batch 64, 8,000 steps. Energies are recovered up to a single global gauge offset (magnitude-NCC’s invariance to per-pair global phase), closed by a one-scalar variational anchor

$$E_0^{\text{calibrated}} = \langle \psi_0^{\text{learned}} | \hat{H} | \psi_0^{\text{learned}} \rangle, \quad \langle \psi | \hat{H} | \psi \rangle = -i \partial_t \langle \psi | U(t) | \psi \rangle |_{t=0},$$

read from the single-frequency oscillation of $\langle \psi_0^{\text{learned}} | U(t) | \psi_0^{\text{learned}} \rangle$ across pair times.

Failure modes at large τ . As τ_{\max} grows past $2\pi/E_{\max}$, the per-pair phase $E_n \tau_b / \hbar$ exceeds the principal branch and the reconstruction loss accepts any wrap count $E_n \tau_b / \hbar + 2\pi m$ as a valid solution. The per-filter rate E_k provides per-filter robustness (each filter settles on a single consistent unwrap), but different filters in the bank can settle on different branches without mutual penalty, producing the spectrum fragmentation in Figure 8 (§4.3). Eigenstate recovery is unaffected: the structural primitive that identifies invariant subspaces (the residual) does not depend on τ .

Position vs. existing literature. Neural Phase Correlation (NPC) recovers the eigenstructure of a unitary operator directly from observation pairs: a single trained model returns the entire spectrum, never sees the Hamiltonian, and does not require the per-pair time interval τ to be observed or binned. Koopman-net learning [21, 22] instead approximates the operator on a learned latent rather than recovering the eigenstructure directly. Neural-quantum-state variational Monte Carlo [23] requires the Hamiltonian and recovers one eigenstate per

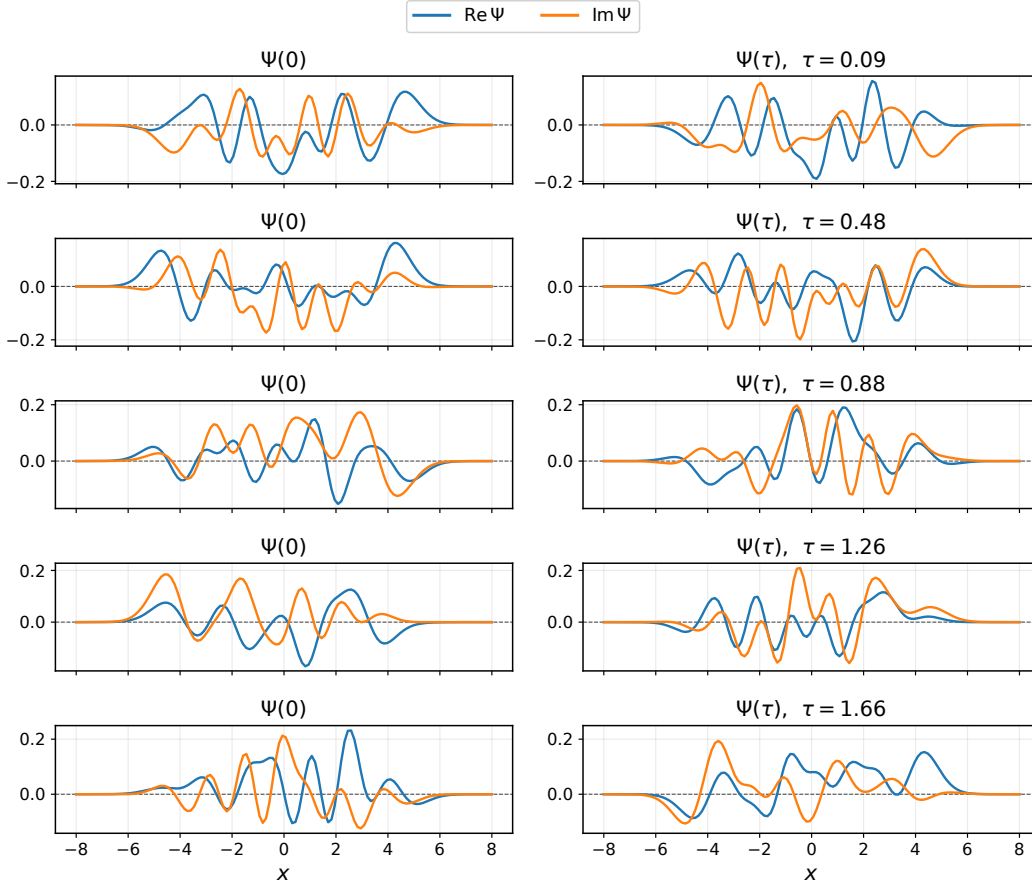


Figure 9: Example training pairs $\Psi(0)$ and $\Psi(\tau)$: random complex unit-norm superpositions of the first $n_{\max} = 16$ Hermite eigenstates, time-evolved analytically by a τ drawn uniformly from $[\tau_{\min}, \tau_{\max}]$. The model receives only the pair, not τ or the coefficients.

training run. Physics-informed eigenvalue learning [24] and Hamiltonian learning via neural ODEs [25] both touch H explicitly during training. Classical dynamic mode decomposition [38] recovers the spectrum from snapshot pairs at a single fixed τ ; the bilinear-interaction residual’s τ -invariance is what lets NPC operate without per-pair τ alignment, where DMD-style methods would require per- τ binning.

E ACDC ablations

E.1 Residual-masking sweep

We ablate the residual-masking mechanism by varying the top- K mask budget at fixed filter-bank capacity $C = 128$, training each variant under our preferred recipe from Appendix D.1 (NCC similarity, diffusion regularization, Jacobian-determinant fold penalty; seed 43, batch size 100 for $C = 128$ and 50 for $C \geq 256$ due to memory, otherwise identical to the Table 2 protocol). The sweep covers $K \in \{1, 8, 32, 64, 96, 128\}$; $K = 128$ recovers the unmasked baseline in which every pair contributes unconditionally. Figure 11 reports Dice as a function of K (left) and the corresponding learned filter banks at each K (right).

Dice peaks Pooled Dice rises from $K = 1$ (0.831) through $K = 64$ (0.857) and then declines as K approaches C , ending at 0.841 unmasked. The drop from peak to unmasked is 0.016, more than $5\times$ the cross-seed standard deviation reported in Appendix D.1. Both directions follow the same shape: ED→ES gains 0.024 from unmasked to peak (0.819 → 0.834), ES→ED gains 0.019 (0.862 → 0.881). Pulling the full bank into the gradient does not just stop helping, it actively hurts, because the high-residual pairs contribute mis-aligned phase signal at inference.

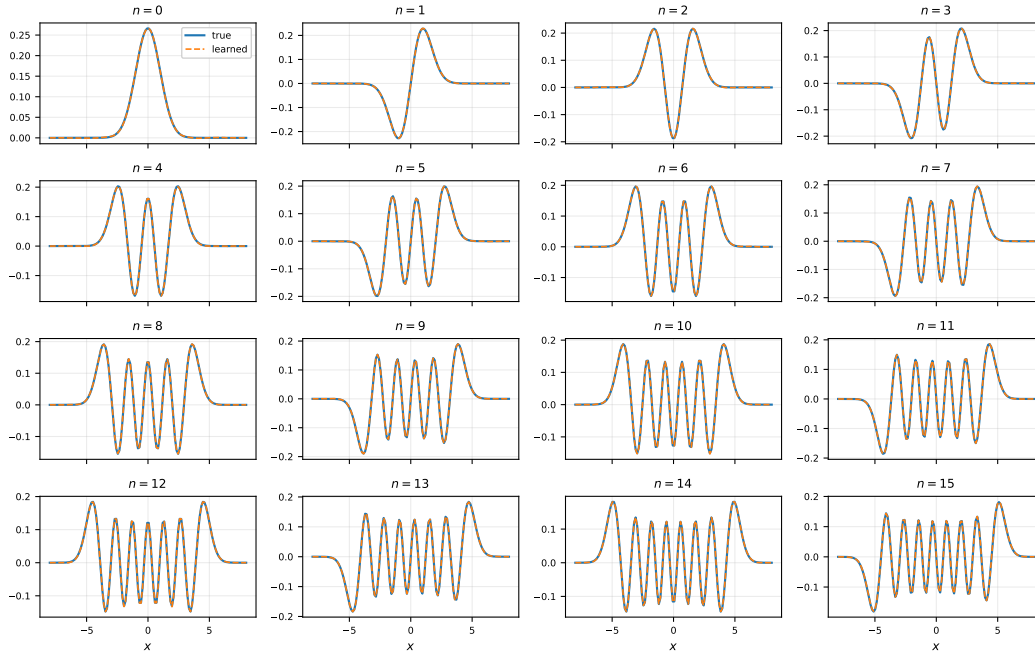


Figure 10: Learned filters (orange dashed) overlaid on the analytic Hermite-function eigenstates (blue) for $n = 0, \dots, 15$, sign-aligned. All 16 filters coincide with the analytic eigenstates (median overlap 1.000).

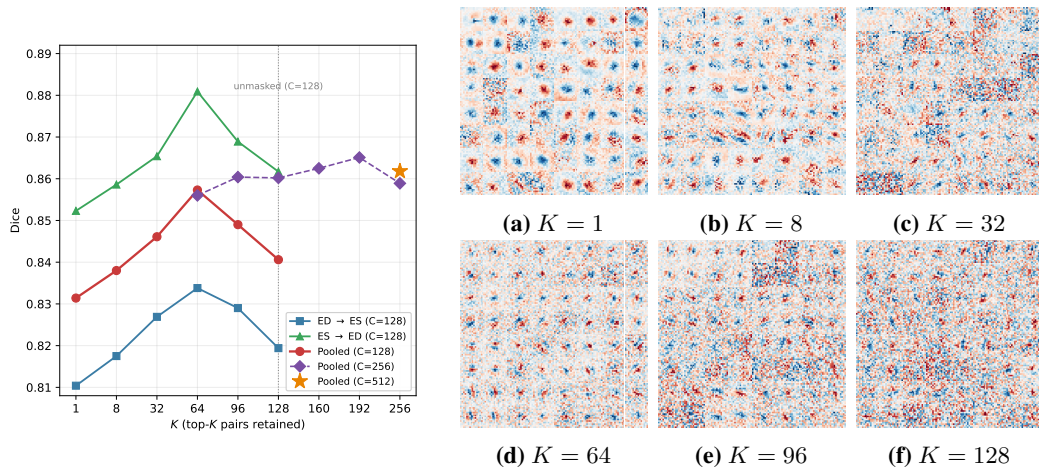


Figure 11: K-sweep ablation on ACDC. The Pooled ($C = 128$) curve forms a smooth dome with peak at $K = 64$ (0.857), 0.016 above the unmasked limit at $K = C = 128$. Purple diamonds overlay Pooled Dice at $C = 256$ (bs=50) for $K \in \{64, 96, 128, 160, 192, 256\}$; the highest Pooled Dice in the sweep is $C = 256$, $K = 192$ (0.865). The orange star marks $C = 512$, $K = 256$ (0.862). **Right:** 16 filter pair quads of each $C = 128$ bank shown as 2×2 quads of $(\psi_{2k-1}, \phi_{2k-1}, \psi_{2k}, \phi_{2k})$. Small K develops clean Gabor-like specialists with strong within-pair coherence; K approaching the unmasked limit produces noise-like generalists with no coherent within-pair structure.

$K = 1$ is the most informative point. A single residual-selected pair already delivers 0.831 Dice, within 0.010 of the unmasked $C = 128$ baseline. Even one specialist captures most of the registration signal, providing direct evidence that the residual is selecting the right pair (the one whose span is most invariant under the local transformation), not merely providing a generic sparsity bottleneck. This connects to the necessity-and-sufficiency result in Appendix B: $r_k^2 = 0$ iff $\text{span}(\psi_k)$ is invariant under L_p , so masking by residual selects pairs with the theoretical guarantee, not a heuristic proxy.

Bank-level diagnostic. The shape transition in Figure 11 (right) mirrors the per-pair residual distribution diagnostic of Figure 3 (main text). The K -sweep makes the supervision-incentive picture from Section 3 explicit: as K shrinks, the supervision concentrates on a tighter specialist set, producing the cleaner Gabor structure

visible in Figure 11 (right). Visual inspection reveals an increase in the noise and granularity of the filter pairs (specialists \rightarrow generalists) for increasing K , except for the apparent lucidity at $K = 64$ which coincides with the sharp Dice peak.

Capacity scaling. Doubling the bank to $C = 256$ leaves the optimal mask budget within the same regime, peaking at $K = 192$ with Pooled Dice 0.865, only 0.008 above the $C = 128$ peak (Figure 11, purple diamonds). Quadrupling to $C = 512$ at $K = 256$ returns 0.862, no better than the unmasked $C = 256$ limit. The peak-then-decay shape persists at $C = 256$, with the $K = 256$ endpoint dipping back to the unmasked level (0.859). Capacity beyond $C = 128$ is therefore not the binding constraint at this benchmark; specialization through a residual-selected subset is.

E.2 ODE step count

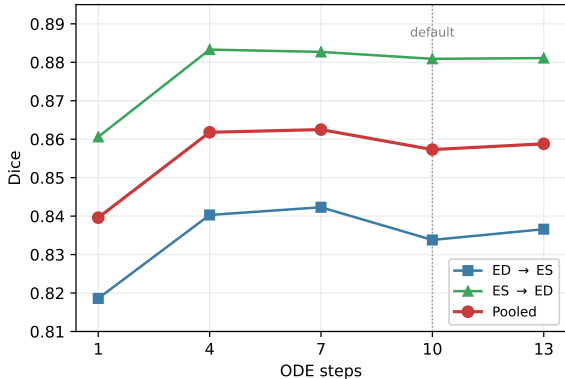


Figure 12: ODE-step ablation on ACDC for the $C=128$, $K=64$ baseline. ODE 10 is the default used throughout the paper (dotted vertical). Pooled Dice plateaus from ODE 4 onward, and the spread between ODE 4 and ODE 13 (0.857 – 0.863) is within the cross-seed standard deviation reported in Appendix D.1.

We sweep the number of integration steps used by the displacement-field ODE while holding all other hyperparameters fixed at the $C = 128$, $K = 64$ baseline of Appendix D.1.

Plateau onset. Pooled Dice rises sharply from a single Euler step (0.840) to ODE 4 (0.862) and then plateaus through ODE 13 (0.859): the spread across the plateau (0.005) sits within the cross-seed standard deviation reported in Appendix D.1. The default of ODE 10 used throughout the paper is inside this plateau, not at its edge.

Inference-cost implication. The plateau means that any step count between 4 and 13 is statistically equivalent on Pooled Dice. ODE 4 attains the same accuracy as the default at roughly $0.4\times$ the per-pair inference cost (Appendix E.3); the choice of ODE 10 in the released checkpoints reflects training-time conservatism rather than an accuracy requirement.

E.3 Compute profile

Setup. We profile the default $C = 128$, $K = 64$, ODE 10 model on a single NVIDIA A100-SXM4-40GB at batch size 1 and input resolution 128×128 , using `fvcore` for FLOPs and 100 forward passes (20 warmup) for wall-clock.

Parameters, FLOPs, and wall clock. The model has 0.32M trainable parameters and 3.29 GFLOPs per forward pass. Mean per-pair inference time is 75.7 ± 1.7 ms (median 75.3 ms, 95th percentile 77.8 ms). Combined with the ODE-step plateau (Appendix E.2), reducing to ODE 4 brings the per-pair cost to approximately 30 ms with no measurable Dice cost; this is the operating point we recommend for inference-throughput-bound deployment.

Training compute. All training runs in the paper were performed on a single NVIDIA A100-SXM4-40GB GPU.

F Deformation breadth

To verify that NPC can operate across a wide diversity of deformation regimes, we report two additional experiments beyond the cardiac MRI and echocardiography benchmarks of Section 4: a qualitative demonstration on 2D OASIS brain MRI (Section F.1) and a quantitative transfer to sub-meter synthetic-aperture-radar imagery (Section F.2), the latter compared against the classical phase-correlation baseline.

F.1 2D OASIS brain MRI

OASIS pairs differ from ACDC in tissue contrast, anatomical structure, and scale of deformation: cortical folds and ventricles dominate the alignment task rather than ventricular contraction.

Setup. We train a single-stage $C = 128$, $K = 64$ model on 2D mid-axial slices from the OASIS-1 dataset [37] using NCC similarity, diffusion regularization, and ODE integration. This section is a qualitative demonstration: we do not compare against OASIS-specific baselines.

Qualitative registration. On four held-out test pairs (Figure 13a), the deformation grid contracts smoothly across ODE steps to align cortical folds and ventricles between moving and fixed slices. The warp tightens monotonically with integration time and remains diffeomorphic at $t = 15$ without visible folding.

Learned filter bank. The learned filter bank shows the same within-pair coherence as on ACDC, with tighter spatial localization (Figure 13b).

F.2 Sub-meter SAR

High-resolution SAR-SAR co-registration is a current bottleneck for downstream remote-sensing tasks (change detection, deformation monitoring, time-series analysis), where sub-pixel alignment is required. The field currently relies on hand-crafted phase-correlation methods; no deep-learning architecture exists, and no public benchmark is available[39]. We therefore construct our own dataset and benchmark against the de facto standard (global phase correlation).

Setup. Our dataset consists of 256×256 geo-coded SAR image pairs at 0.5 m resolution from locations throughout Europe, Asia, Australia, North America, and South America. We present results on held-out test pairs taken over the Puerto Maldonado region of the Peruvian Amazon.

Registration results. Across the 40 sampled pairs (Figure 14a), the network beats classical phase correlation on all but one pair, which we show in the figure, demonstrating that the model transfers to phase correlation’s operational domain without framework modifications.

Cascade filter bank. The coarse-stage filters (Figure 14b, top two rows) develop broader, speckle-tolerant patterns; the fine-stage filters (bottom two rows) tighten to higher spatial frequencies, consistent with the coarse-to-fine training schedule.

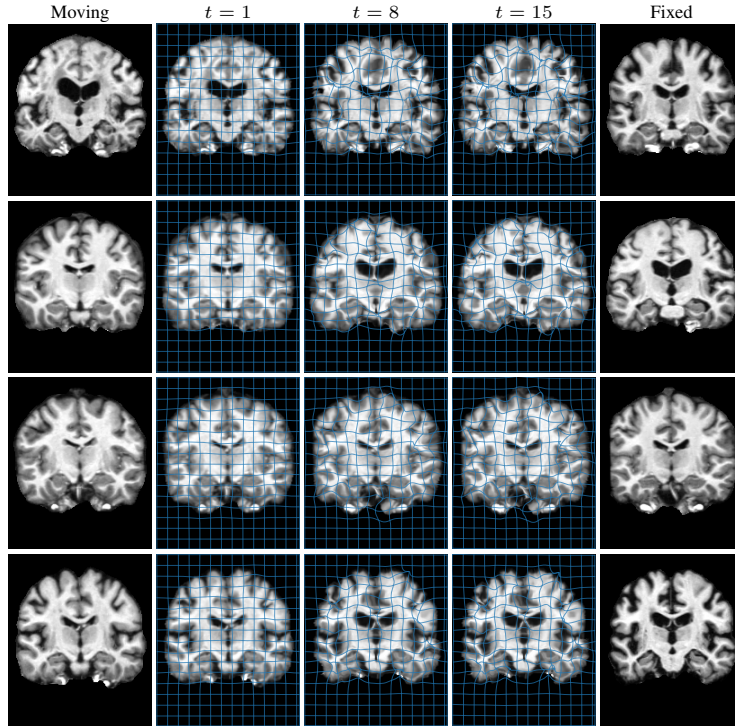
F.3 Filter localization across regimes

The three regimes span a spectrum of filter localization. SAR filters are not spatially localized (Figure 14b). ACDC filters concentrate moderately, on the scale of anatomical features (Figure 11). OASIS filters are the most localized, with effective half-periods of approximately 4 pixels, at the spatial scale of individual cortical features (Figure 13b).

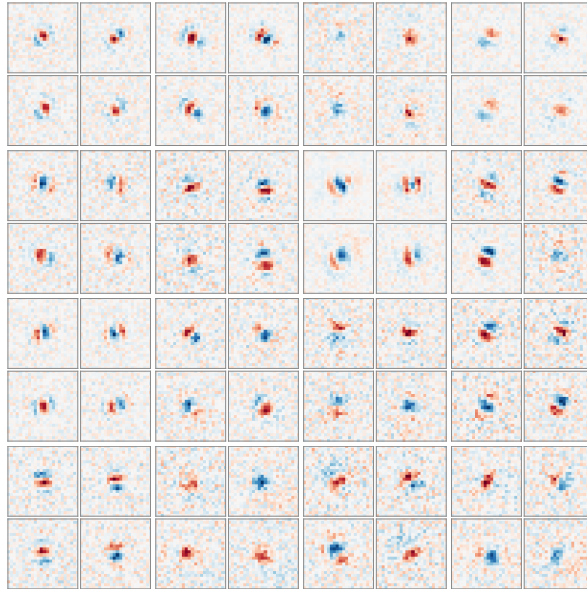
G Limitations

The framework assumes the local transformation L_p is approximately orthogonal: locally, on small patches, for image registration; exactly for unitary dynamics. Some classes of orthogonal transformations are not captured by a trivial extension of the current architecture. In particular, optical flow with non-local orthogonal structure (transformations that mix pixels at distant locations) requires a non-local cost-volume formulation rather than the per-location bilinear interaction used here.

For the QHO experiment, we use only the first 16 Hermite eigenstates with a matched bank of 16 filter pairs; the architecture’s behavior at higher truncation orders is not characterized. The reconstruction loss is invariant to per-pair global phase, which causes filters to fragment across unwrap branches at large τ_{\max} (Appendix D.3) and limits the model’s ability to recover the eigenenergies (and therefore the Hamiltonian) at large time intervals.



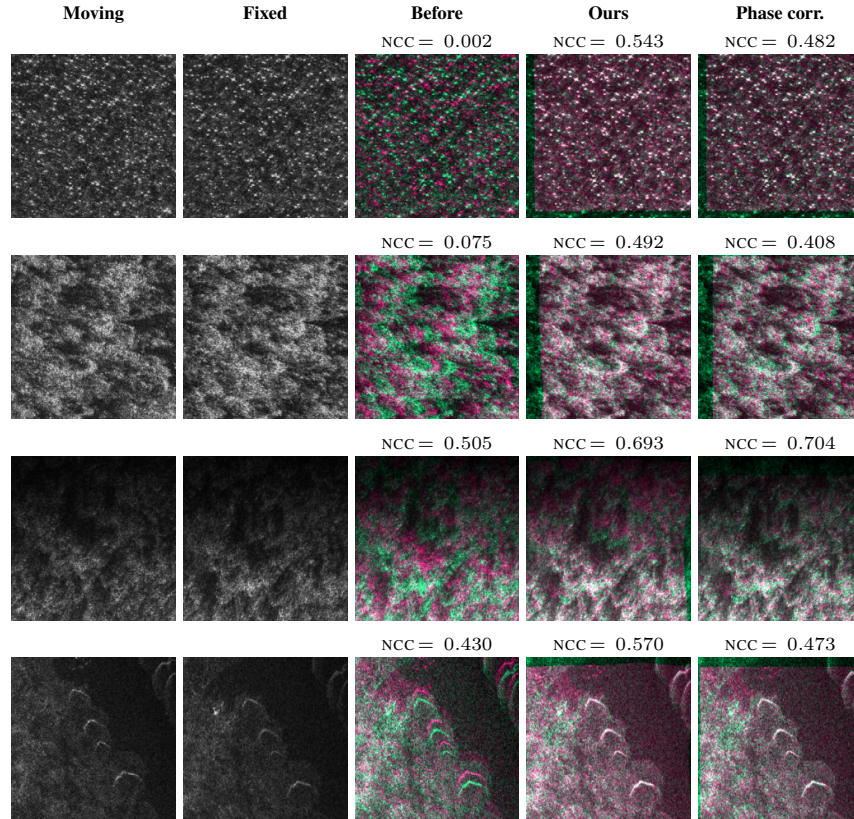
(a) Registration timelapse on four held-out test pairs. Columns: moving, intermediate warps at ODE steps $t \in \{1, 8, 15\}$ with deformation grid overlay, and fixed.



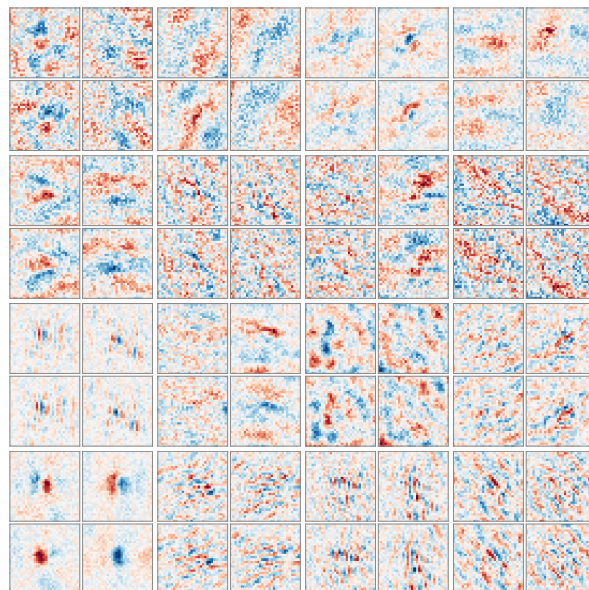
(b) Sixteen learned quad filter pairs from the OASIS bank. Each panel is a 2×2 quad of $(\psi_{2k-1}, \phi_{2k-1}, \psi_{2k}, \phi_{2k})$.

Figure 13: OASIS 2D cross-domain demonstration. (a) Deformation grid contracts smoothly across ODE steps to align cortical folds and ventricles, remaining diffeomorphic at $t = 15$ without visible folding. (b) Learned filter bank exhibits the same spatially localized, within-pair coherent structure observed on ACDC (Figure 11, right).

The QHO experiment is also a purely algebraic exercise: training pairs are generated by analytic time-evolution of complex superpositions. The setup is not physically realizable in its current form because quantum-state measurements are destructive (wavefunction collapse), so a physical experiment cannot directly observe $\Psi(0)$ and $\Psi(\tau)$ as the model requires.



(a) SAR-to-SAR registration on Puerto Maldonado Umbra acquisitions (Feb24 → Mar20). Network NCC exceeds phase-correlation NCC on all but one sample (row 3, included to show the failure case).



(b) Sixteen learned quad filter pairs from the cascade: coarse stage (top two rows) and fine stage (bottom two rows). Each panel is a 2×2 quad of $(\psi_{2k-1}, \phi_{2k-1}, \psi_{2k}, \phi_{2k})$.

Figure 14: SAR cross-domain transfer. (a) Registration results on four held-out Umbra image pairs. (b) Coarse- and fine-stage learned filter banks. The coarse stage develops broader, speckle-tolerant patterns; the fine stage tightens to higher spatial frequencies.