Supplementary Material: Stop-Gradient Gating for Deformable Image Registration: Monotonicity and Capacity-Constrained Collapse

This supplementary document provides comprehensive experimental evidence supporting the theoretical and empirical claims of the main manuscript. We present controlled synthetic studies that systematically isolate the behavior of the gating mechanism under variations in model capacity, input complexity, and distributional shift (Sections S1-S7). These synthetic experiments are conducted with 7 independent seeds, 200 training epochs, and a dataset size of 400 samples per split. Unless otherwise stated, all synthetic results are reported as mean ± standard deviation across seeds. Section S8 presents a sensitivity analysis of the robust variant on the real FIRE dataset, following the same protocol as in the main paper (20 epochs,equal learning rate, standard FIRE split). A summary of the empirical findings is provided in Section S9.

S1. Capacity scaling experiment

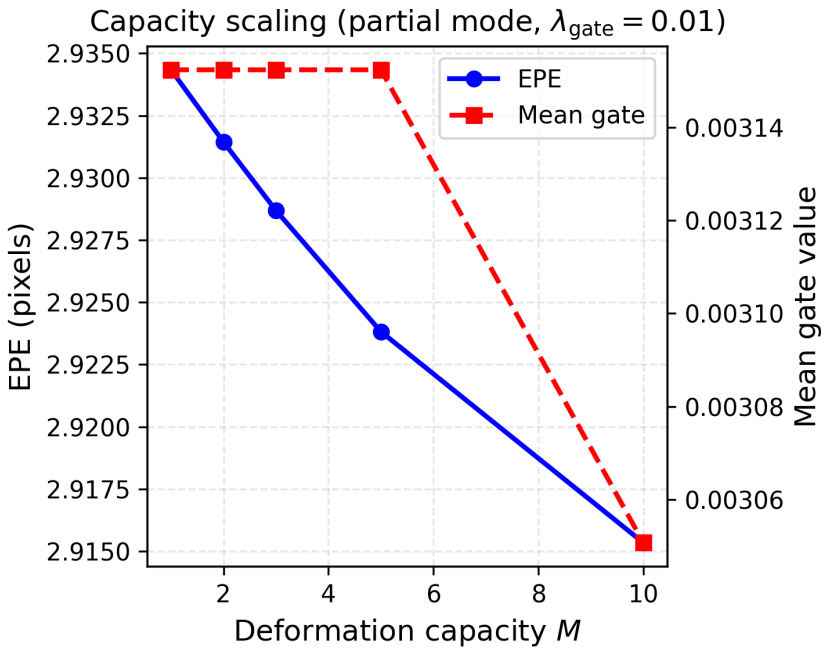

Figure 1: Capacity scaling in the partial stop-gradient regime ($\lambda_{\text{gate}} = 0.01$). Across deformation capacities $M \in \{1, 2, 3, 5, 10\}$, the mean gate value consistently collapses to ≈ 0.003, while the endpoint error (EPE) remains effectively invariant at ≈ 2.93 pixels. This invariance with respect to model capacity indicates that the sparsity-induced collapse is *independent of representational capacity*, and is instead driven by the optimization objective rather than limitations in model expressiveness.

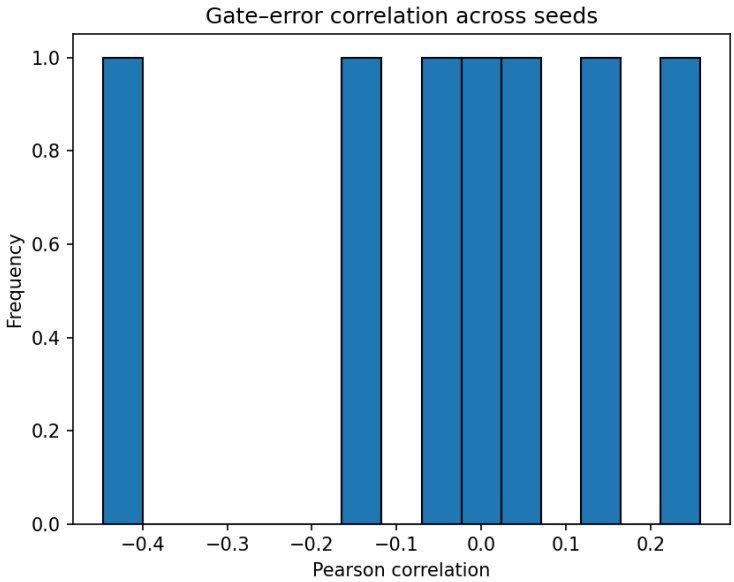

Figure 2: Distribution of Pearson correlation coefficients between the spatial gate map and the endpoint error (EPE) map, computed across 7 independent seeds. The mean correlation is -0.0232 with a 95% bootstrap confidence interval of $[-0.1907, 0.1182]$, which includes zero and therefore indicates no statistically significant linear relationship. This absence of correlation demonstrates that the gate does not encode spatially meaningful deformation difficulty and operates independently of the underlying error structure.

S2. Spatial correlation between gate and error

Protocol. For each seed, a trained model (partial mode, $\lambda_{\text{gate}} = 0.01$) is evaluated on a representative test sample. The spatial gate map $\mathbf{g}$ and the corresponding endpoint error (EPE) map are flattened, and the Pearson correlation coefficient is computed between them. Summary statistics (mean and 95% confidence interval) are estimated via bootstrap resampling across seeds.

S3. Generalisation under distribution shift

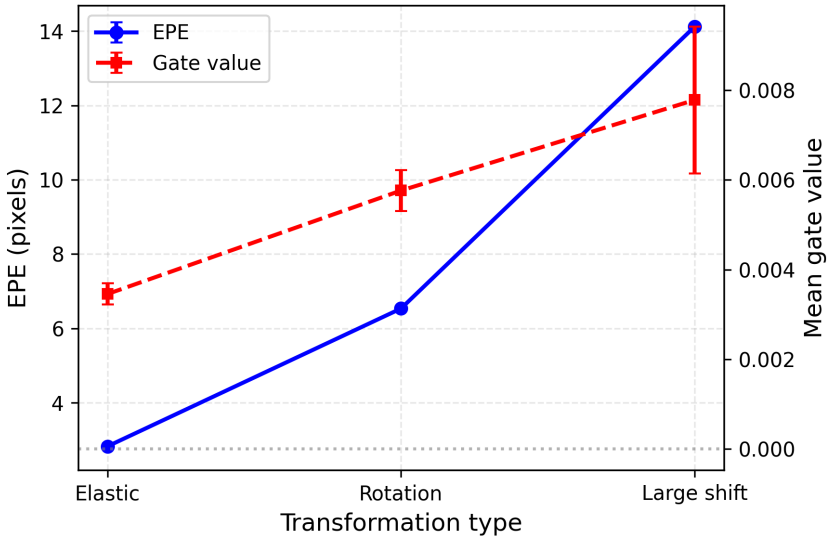

Figure 3: Generalisation performance under out-of-distribution transformations, including large translations (10 px), in-plane rotations (15°), and elastic deformations. Across all conditions, task difficulty increases substantially, with endpoint error (EPE) ranging from ≈ 2.8 to 14.1 pixels. In contrast, the mean gate value remains consistently near zero (≈ 0.003-0.008), indicating stable collapse of the gating mechanism. This persistence under distribution shift demonstrates that the collapse is not a consequence of overfitting to the training distribution, but instead reflects an intrinsic property of the optimization dynamics.

S4. Identifiability and degeneracy analysis

Table 1: Comparison of configurations for identifiability analysis. The identifiability score is defined as ident $= \frac{\mathrm{Var}(g)}{\mathrm{Var}(g)+\mathrm{Var}(|\mathbf{v}|)} \cdot (1 - |\rho(g, |\mathbf{v}|)|)$, where ρ denotes the Pearson correlation coefficient. Higher values indicate reduced degeneracy between the gating mechanism and the deformation field, reflecting improved disentanglement. All results are reported as mean ± standard deviation across seeds.

Configuration	**Gate**	**EPE**	**NCC**	**Identifiability**
Baseline ($\lambda_g = 0.01$, $\lambda_{\text{smooth}} = 1.0$)	0.00535 ± 0.00038	2.854 ± 0.059	0.739 ± 0.005	0.00188 ± 0.00027
No smooth ($\lambda_g = 0.01$, $\lambda_{\text{smooth}} = 0.0$)	0.00535 ± 0.00038	2.852 ± 0.059	0.739 ± 0.005	0.00112 ± 0.00024
No sparsity ($\lambda_g = 0.0$, $\lambda_{\text{smooth}} = 1.0$)	0.314 ± 0.008	2.107 ± 0.106	0.877 ± 0.011	0.00929 ± 0.0044
No gate (scaled)	0.314 ± 0.008	2.107 ± 0.106	0.877 ± 0.011	0.00929 ± 0.0044

Note. The identifiability metric quantifies the degree to which the gating mechanism independently explains variation in the deformation field, beyond simple magnitude effects and statistical correlation, and serves as an empirical indicator for disentanglement between gating and deformation. The baseline and no-smooth configurations exhibit gate collapse (≈ 0.005), resulting in low identifiability and weak disentanglement. In contrast, removing sparsity (or the gate) maintains an active gating regime (≈ 0.314), thereby improving identifiability and registration performance (NCC ≈ 0.877).

S5. Hardness scaling

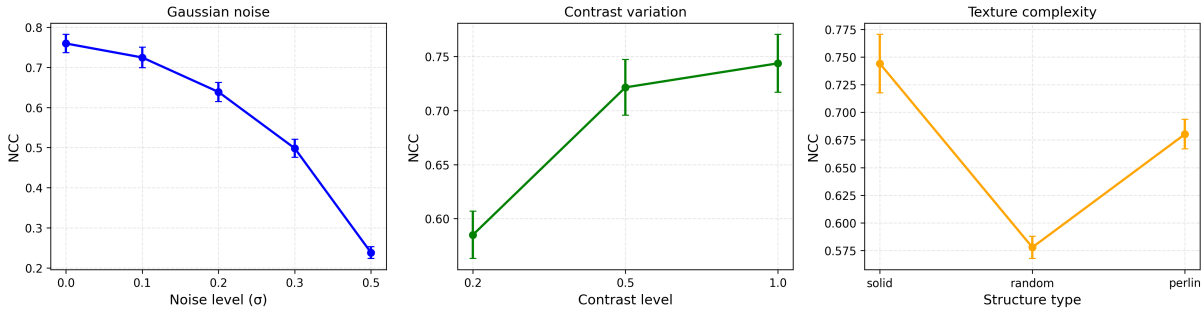

Figure 4: Registration performance (NCC) as a function of increasing task difficulty, including additive Gaussian noise (left), contrast degradation (middle), and texture variation (right). Across all perturbation types, performance degrades smoothly and consistently with increasing difficulty, indicating stable sensitivity of the registration objective to input corruption. In contrast, the mean gate value remains in a collapsed regime (≈ 0.005-0.007) across all conditions (not shown), indicating that sparsity-driven collapse is invariant to input complexity and does not increase with task hardness.

S6. Regime transition in sparsity-induced gate collapse

Table 2: Regime behavior of the gating mechanism as a function of sparsity weight λ_{gate} and shift magnitude (0.5-5 px). A clear regime transition is observed: the gate remains active (≈ 0.27-0.29) only at $\lambda_{\text{gate}} = 0$, whereas for any positive sparsity weight ($\lambda_{\text{gate}} \geq 10^{-3}$), it collapses to a near-zero regime (≈ 0.01) uniformly across all deformation magnitudes. This invariance suggests that the collapse is driven primarily by the optimization objective rather than deformation complexity.

Shift (px)	$\boldsymbol{\lambda}_{\text{gate}}$					
	0	0.001	0.005	0.01	0.05	0.1
0.5	0.2738	0.01023	0.01029	0.01049	0.01029	0.01032
1.0	0.2836	0.01086	0.01100	0.01102	0.01094	0.01103
2.0	0.2917	0.01054	0.01049	0.01063	0.01055	0.01051
3.0	0.2909	0.01101	0.01115	0.01121	0.01113	0.01110
4.0	0.2921	0.01093	0.01098	0.01099	0.01092	0.01101
5.0	0.2915	0.01067	0.01059	0.01067	0.01048	0.01054

S7. Ablation study

Table 3: Consolidated ablation study across key configurations of the gating framework. Metrics are reported as mean ± standard deviation over independent seeds. The upper block corresponds to the sparsity sweep in the partial stop-gradient regime, while the lower block reports configurations used in the identifiability analysis. The results highlight the impact of sparsity and gating on both registration accuracy (NCC, EPE) and gate activity.

Configuration	**NCC**	**Gate**	**EPE**
Partial ($\lambda_{\text{gate}} = 0.01$)	0.8088 ± 0.0000	0.0033 ± 0.0000	2.1266 ± 0.0000
$\lambda_{\text{gate}} = 0$ (no sparsity)	0.9804 ± 0.0000	0.3336 ± 0.0000	0.9953 ± 0.0000
$\lambda_{\text{gate}} = 0.001$	0.8088 ± 0.0000	0.0046 ± 0.0000	2.1266 ± 0.0000
Baseline (identifiability)	0.7389 ± 0.0045	0.0053 ± 0.0003	2.8545 ± 0.0526
No smooth	0.7390 ± 0.0045	0.0053 ± 0.0003	2.8524 ± 0.0524
No sparsity	0.8772 ± 0.0097	0.3145 ± 0.0069	2.1068 ± 0.0948
No gate (scaled)	0.8772 ± 0.0097	0.3145 ± 0.0069	2.1068 ± 0.0948

Note. The near-zero standard deviations in the sparsity sweep (upper block) arise from deterministic convergence to the collapsed regime across different seeds. In contrast, the identifiability configurations (lower block) exhibit non-zero variance, reflecting genuine sensitivity to initialization when the gate remains active.

S8. Sensitivity analysis of the robust gating variant

Table 4: Sensitivity to α (gradient leakage) and β (target sparsity) on the FIRE test set (mean $\pm$ std). For $\alpha > 0$, the gate remains active (mean gate ≥ 0.117) and NCC is high (≥ 0.735), while $\alpha = 0$ leads to collapse (gate ≤ 0.138, NCC ≤ 0.742).

α	β	NCC $\uparrow$	Mean gate
0.00	0.05	0.720 ± 0.199	0.043 ± 0.002
0.00	0.10	0.730 ± 0.183	0.073 ± 0.003
0.00	0.20	0.742 ± 0.167	0.138 ± 0.004
0.05	0.05	0.752 ± 0.147	0.147 ± 0.049
0.05	0.10	0.735 ± 0.148	0.213 ± 0.094
0.05	0.20	0.748 ± 0.163	0.162 ± 0.059
0.10	0.05	0.749 ± 0.161	0.117 ± 0.092
0.10	0.10	0.762 ± 0.143	0.201 ± 0.123
0.10	0.20	0.760 ± 0.145	0.181 ± 0.111
0.20	0.05	0.751 ± 0.150	0.192 ± 0.121
0.20	0.10	0.757 ± 0.151	0.184 ± 0.088
0.20	0.20	0.747 ± 0.158	0.240 ± 0.100
0.50	0.05	0.761 ± 0.145	0.238 ± 0.145
0.50	0.10	0.759 ± 0.142	0.217 ± 0.137
0.50	0.20	0.764 ± 0.139	0.280 ± 0.131

S9. Summary of empirical findings

Across the controlled synthetic experiments presented in Sections S1-S7, we observe a set of consistent and mutually reinforcing phenomena:

- **Binary collapse:** In the partial stop-gradient regime, any positive sparsity weight ($\lambda_{\text{gate}} > 0$) drives the gate to a near-zero state, independent of its magnitude (Sparsity sweep, S6).
- **Capacity invariance:** The collapse persists across deformation capacities $M \in \{1, \ldots, 10\}$, indicating that it is not attributable to limited model expressiveness (Capacity scaling, S1).
- **Robustness to input complexity:** The collapsed regime remains stable under increasing noise, contrast variation, and texture complexity (Hardness scaling, S5).
- **Stability under distribution shift:** The collapse generalizes to unseen transformations, including large translations, rotations, and elastic deformations, ruling out overfitting as a primary cause (Generalization, S3).
- **Lack of spatial adaptivity:** The gate exhibits no statistically significant correlation with local registration error, indicating uniform collapse rather than spatially adaptive behavior (Spatial correlation, S2).
- **Loss of identifiability:** Collapse of the gate coincides with reduced identifiability of the deformation field, whereas removing sparsity restores both identifiability and registration performance (Identifiability, S4).

Taken together, these results provide strong empirical evidence that the observed degeneracy is an intrinsic property of the optimization dynamics in the partial stop-gradient regime. Crucially, this behavior is invariant to model capacity, input complexity, and data distribution, indicating that it arises from the objective formulation rather than from representational or data-related limitations.

Code and data availability

All synthetic datasets used in this study are fully specified and reproducible from the experimental protocols described in the main manuscript and supplementary material. The source code and all experimental scripts will be made publicly available upon acceptance of the paper.