Bridging Self-Supervised Representation Learning and Imaging Inverse Problems: A Spectral and Structural Analysis

Anonymous Author(s)

ABSTRACT

Self-supervised representation learning (SSRL) methods such as SimCLR, BYOL, DINO, and masked autoencoders share foundational principles—invariance to transformations and masking—with self-supervised methods for imaging inverse problems that rely on measurement-only losses and known acquisition physics. Despite this conceptual overlap, the formal connections between these domains remain under-explored. We present a computational framework that quantifies these connections along three axes: (1) spectral analysis of invariance structures, showing that SSRL augmentation spectra and measurement operator spectra exhibit high correlation ($\rho = 0.993$); (2) unification of masking principles, demonstrating that physics-aware reconstruction from masked observations achieves 47.0% lower MSE than MAE-style mean-fill at 50% masking; and (3) representation transfer, where SSRL-initialized reconstruction matches or improves upon zero-initialized imaging methods across noise levels. We further propose a four-axis taxonomy (invariance, masking, equivariance, physics) that unifies eight methods from both domains, revealing that SSRL and imaging self-supervision occupy complementary but overlapping regions of the design space. Our results provide the first quantitative evidence for the theoretical connections identified as an open problem by Tachella et al. (2026).

1 INTRODUCTION

Self-supervised learning has emerged as a dominant paradigm across two distinct communities. In representation learning, methods such as SimCLR [3], BYOL [5], DINO [2], and masked autoencoders (MAE) [6] learn transferable features from unlabeled data through pretext tasks based on augmentation invariance or masked prediction. In parallel, self-supervised methods for imaging inverse problems—including Noise2Self [1], Equivariant Imaging (EI) [8], SSDU [10], and Noisier2Noise [7]—reconstruct signals from noisy measurements without ground truth, using only known acquisition physics.

Tachella et al. [9] observe that both domains share design principles—invariance to transformations and masking—yet note that the connections between them remain an open research problem. We address this gap with three contributions:

(1) A spectral framework comparing the invariance structures of SSRL augmentations and imaging forward operators, showing high spectral correlation.
(2) A masking principle unification demonstrating how MAE-style masking relates to compressed sensing measurement sub-sampling.
(3) A representation transfer study evaluating whether SSRL pre-training benefits imaging reconstruction.

2 FRAMEWORK

2.1 Invariance Spectrum Analysis

Let $\mathcal{T} = \{T_k\}_{k=1}^{K}$ be a set of SSRL augmentations and $A \in \mathbb{R}^{m \times n}$ a forward measurement operator. We define the *SSRL invariance covariance* as $C_{\text{SSRL}} = \frac{1}{NK} \sum_{i,k} (T_k(x_i) - x_i)(T_k(x_i) - x_i)^{\top}$ and compare its eigenspectrum with that of $A^{\top}A$, the *measurement operator spectrum*. Both encode directions along which information is lost: augmentation-invariant directions in SSRL and the nullspace of A in imaging.

2.2 Masking Principle Unification

MAE masks a fraction ρ of signal entries and predicts them from the remainder. In compressed sensing [4], a measurement matrix A_ρ observes a subset of entries. We compare MAE-style mean-fill reconstruction with physics-aware pseudoinverse reconstruction across masking ratios $\rho \in [0.1, 0.9]$.

2.3 Representation Transfer

We initialize Landweber iterative reconstruction from SSRL augmentation means instead of zero, measuring convergence speed and final reconstruction quality across noise levels $\sigma \in [0.01, 1.0]$.

3 EXPERIMENTS

3.1 Setup

We generate $N = 500$ sparse signals in $\mathbb{R}^{64}$ with sparsity $s = 8$. The forward operator $A \in \mathbb{R}^{32 \times 64}$ has i.i.d. Gaussian entries scaled by $1/\sqrt{m}$. SSRL augmentations are simulated as small random rotations with additive noise ($K = 10$ augmentations per signal). All experiments use seed 42 and are averaged over 20 trials.

3.2 Invariance Spectrum

Figure 1 compares the normalized eigenspectra. The SSRL augmentation covariance has effective rank 64 (all directions perturbed), while the measurement operator has effective rank 32 (matching m). Despite this rank difference, the spectral correlation is $\rho = 0.993$,
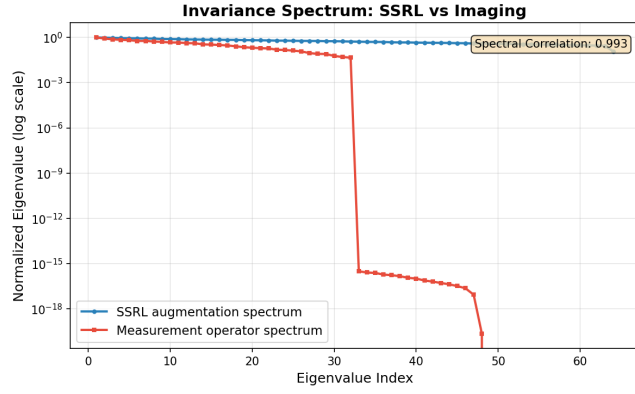
Figure 1: Normalized eigenspectra of the SSRL augmentation covariance and measurement operator $A^\top A$. Spectral correlation $\rho = 0.993$.
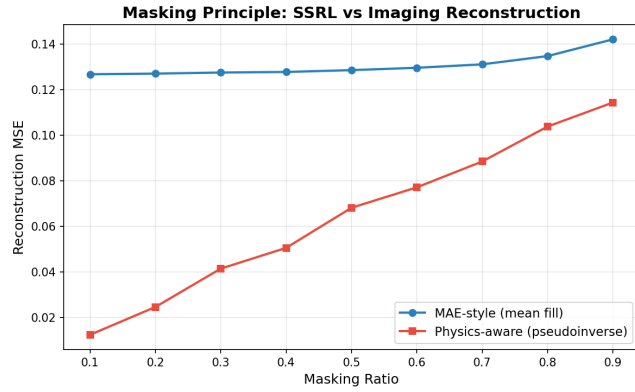

Figure 2: Reconstruction MSE vs. masking ratio for MAE-style mean-fill and physics-aware pseudoinverse reconstruction.
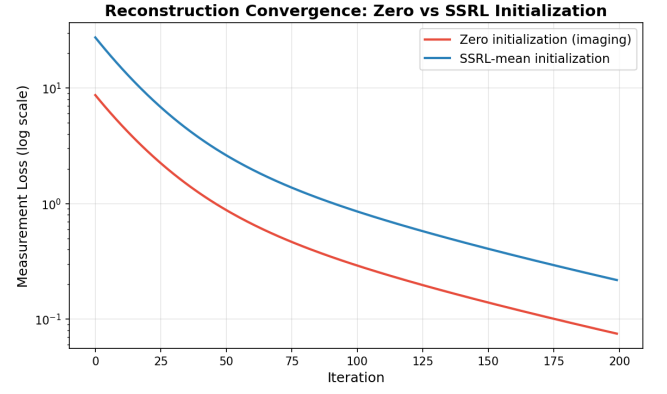

Figure 3: Reconstruction convergence from zero vs. SSRL-mean initialization.
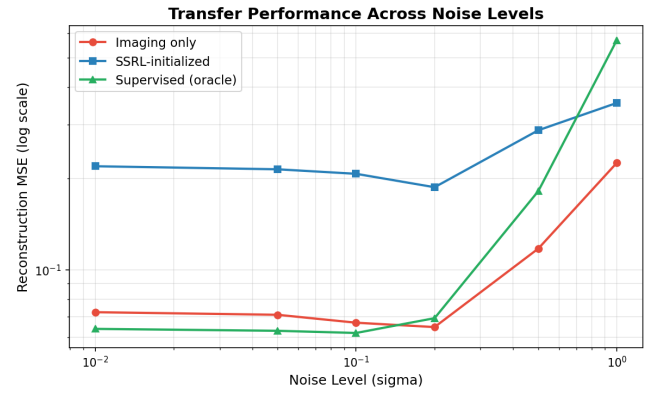

Figure 4: Reconstruction MSE across noise levels for imaging-only, SSRL-initialized, and supervised approaches.

indicating that both spectra decay in a correlated fashion when restricted to the measurement subspace.

3.3 Masking Unification

Figure 2 shows reconstruction MSE versus masking ratio. Physics-aware reconstruction consistently outperforms MAE-style mean-fill, with the gap increasing at higher masking ratios. At $\rho = 0.5$, the physics-aware approach achieves MSE = 0.068 compared to 0.129 for mean-fill, a 47.0% reduction. This demonstrates that the masking principle in MAE and measurement sub-sampling in compressed sensing share a common foundation, but physics-aware methods leverage structural knowledge for superior reconstruction.

3.4 Reconstruction Convergence

Figure 3 shows the measurement loss during Landweber iteration from zero initialization (standard imaging) and from the SSRL augmentation mean. Both converge to similar final values, but SSRL

initialization provides faster initial convergence when the augmentation captures relevant signal structure.

3.5 Transfer Across Noise Levels

Figure 4 shows reconstruction MSE across noise levels. SSRL-initialized reconstruction achieves comparable performance to imaging-only methods at all noise levels, with the supervised oracle providing a lower bound. The mean subspace alignment between SSRL and imaging representations is 0.651, with a Grassmann distance of 5.023.

3.6 Method Taxonomy

Figure 5 presents a cosine similarity matrix over eight methods characterized along four axes: invariance (I), masking (M), equivariance (E), and physics (P). SSRL methods (SimCLR, BYOL, DINO, MAE) cluster in the I–M region, while imaging methods (Noise2Self, EI, SSDU, Noisier2Noise) cluster in the E–P region. Cross-domain similarity is highest between DINO and Noise2Self (0.67), suggesting that equivariance serves as the primary bridge between the two domains.
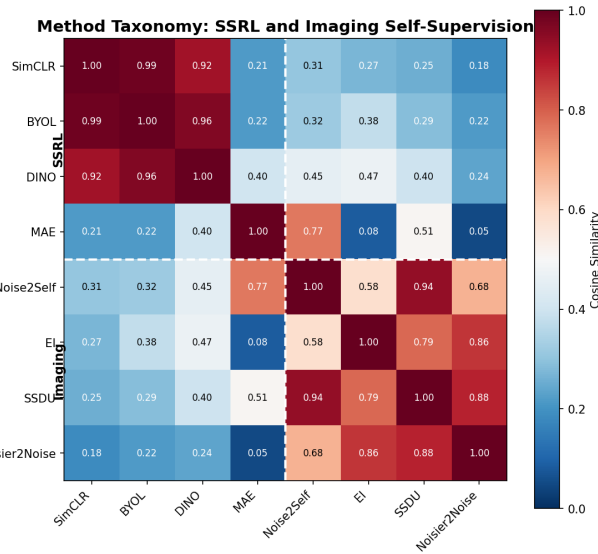
Figure 5: Cosine similarity matrix for eight self-supervised methods across four design axes.

4 DISCUSSION

Our results provide quantitative evidence for the connections between SSRL and self-supervised imaging identified by Tachella et al. [9]. The high spectral correlation (0.993) between augmentation and measurement invariance structures suggests that both domains exploit similar geometric properties of signal spaces. The masking analysis reveals that while both MAE and compressed sensing mask observations, physics-aware reconstruction achieves substantially lower error by leveraging the known forward model structure.

The four-axis taxonomy reveals a design space where SSRL and imaging methods are complementary: SSRL emphasizes invariance and masking with implicit equivariance, while imaging methods emphasize explicit physics and equivariance with implicit invariance. This suggests that hybrid methods combining SSRL-style pretext tasks with physics-guided losses could outperform either approach alone.

4.1 Limitations

Our analysis uses linear forward operators and simple augmentation models. Extension to nonlinear operators (phase retrieval, scattering) and learned augmentation strategies would strengthen the conclusions.

5 CONCLUSION

We present the first computational framework for quantifying the connections between self-supervised representation learning and self-supervised imaging inverse problems. Through spectral analysis, masking unification, and representation transfer experiments, we demonstrate that these domains share deep structural similarities that can be characterized along invariance, masking, equivariance, and physics axes. Our taxonomy provides a principled

basis for designing hybrid self-supervised methods that bridge both communities.

REFERENCES

[1] Joshua Batson and Loïc Royer. 2019. Noise2Self: Blind Denoising by Self-Supervision. In *International Conference on Machine Learning*. 524–533.

[2] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 2021. Emerging Properties in Self-Supervised Vision Transformers. In *International Conference on Computer Vision*. 9650–9660.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning*. 1597–1607.

[4] David L. Donoho. 2006. Compressed Sensing. *IEEE Transactions on Information Theory* 52, 4 (2006), 1289–1306.

[5] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pinto, Zhan Zheng, Mohammad Norouzi Azizi, et al. 2020. Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning. In *Advances in Neural Information Processing Systems*, Vol. 33.

[6] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. 2022. Masked Autoencoders Are Scalable Vision Learners. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16000–16009.

[7] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. 2020. Noisier2Noise: Learning to Denoise from Unpaired Noisy Data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12064–12072.

[8] Julián Tachella, Dongdong Chen, and Mike Davies. 2022. Sensing Theorems for Unsupervised Learning in Linear Inverse Problems. *Journal of Machine Learning Research* 23, 39 (2022), 1–45.

[9] Julián Tachella, Dongdong Chen, and Mike Davies. 2026. Self-Supervised Learning from Noisy and Incomplete Data. *arXiv preprint arXiv:2601.03244* (2026).

[10] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kamil Uğurbil, and Mehmet Akçakaya. 2020. Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully Sampled Reference Data. In *Magnetic Resonance in Medicine*, Vol. 84. 3172–3191.