

Multi-modal self-supervised pre-training of crystallographic representations for data-efficient materials property prediction

Danial Ebrahimzadeh, Sarah S. Sharif, and Yaser Banad*

School of Electrical and Computer Engineering, University of Oklahoma, Norman, Oklahoma 73019, USA

*Corresponding author: bana@ou.edu

S1. Bibliometric overview of vanadium–oxide research (2000–2025)

To contextualize the scientific landscape that motivates the present generative framework, we surveyed the Web of Science Core Collection for publications addressing individual vanadium–oxide compositions over the period 2000–2025. Figure S1 reports the total record counts per composition on a logarithmic scale, revealing a strongly skewed distribution in which V_2O_5 (17,497 records) and VO_2 (8,773 records) account for the majority of the literature, followed by V_2O_3 (2,222) and VO (1,090). Reduced Magnéli-series and intermediate-valence phases such as V_6O_{13} , V_3O_7 , V_3O_5 , V_4O_9 , V_4O_7 , V_6O_{11} , V_5O_9 , V_7O_{13} , V_8O_{15} , and V_9O_{17} remain comparatively under-represented, with record counts ranging from a few hundred to fewer than ten. Figure S2 shows that the dominant phases, V_2O_5 and VO_2 , have undergone sustained and near-exponential growth in annual publication output, particularly after 2012, consistent with the resurgence of interest in thermochromic infrared modulation, multivalent-ion batteries, electrochromic coatings, and Mott-transition-based neuromorphic devices. Figure S3 demonstrates that the lower-frequency phases, although still small in absolute terms, exhibit clear upward trends over the same period, indicating a broadening of research attention beyond the conventional VO_2 – V_2O_5 axis toward less-explored stoichiometries with potentially distinctive electronic and electrochemical functionalities. Figure S4 arranges the most highly cited publications for the five most-studied compositions (V_2O_5 , VO_2 , V_2O_3 , VO , V_6O_{13}) along a publication-year timeline, with box colors matched to the composition palette used in Figure S2. The papers identified by this analysis collectively span thermochromic device physics, infrared and Mott-transition spectroscopy, electrochemical and electrochromic energy storage, heterogeneous catalysis, and correlated-electron behavior, illustrating both the breadth of V–O applications and the system's sustained scientific impact. Together, the four panels establish that the V–O system is simultaneously large in volume, diversifying in composition, and sustained in scientific impact, an empirical motivation for inverse-design frameworks capable of generating candidate structures across the full binary compositional landscape rather than within a few well-characterized phases.

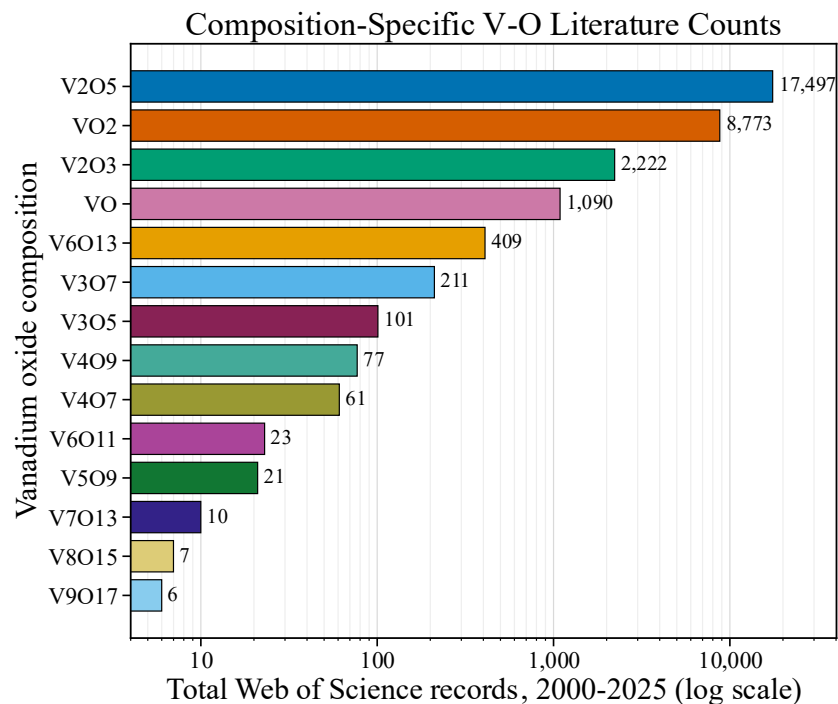


Figure S1: Total Web of Science records (2000-2025) for vanadium-oxide compositions, displayed on a logarithmic scale. V_2O_5 (17497 records) and VO_2 (8773) dominate the literature, while reduced Magneli-series phases such as V_7O_{13} , V_8O_{15} , and V_9O_{17} remain comparatively under-represented

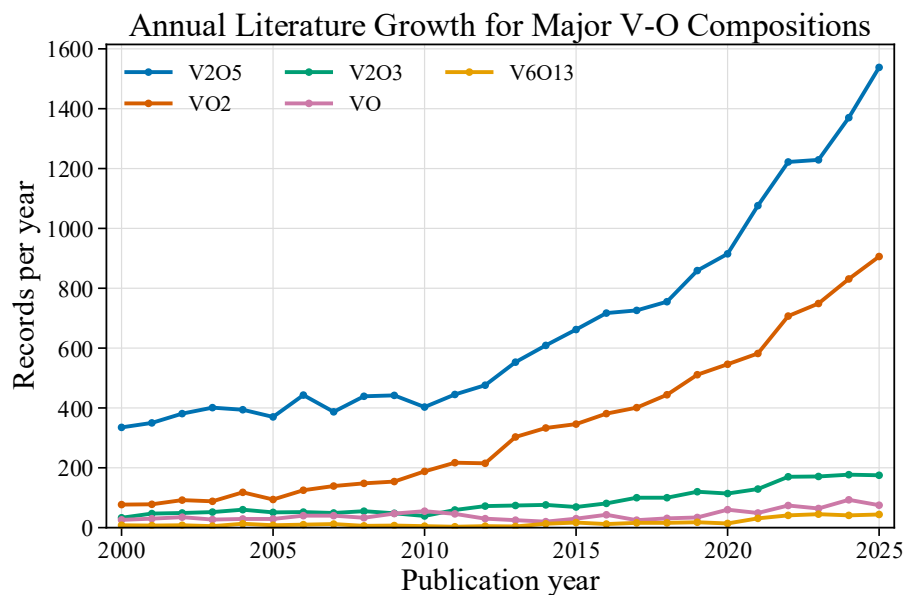


Figure S2: Annual publication counts (2000-2025) for the five most-studied V-O compositions. V_2O_5 and VO_2 exhibit sustained, near-exponential growth, particularly after 2012, reflecting renewed interest driven by energy-storage, electrochromic, thermochromic, and neuromorphic applications

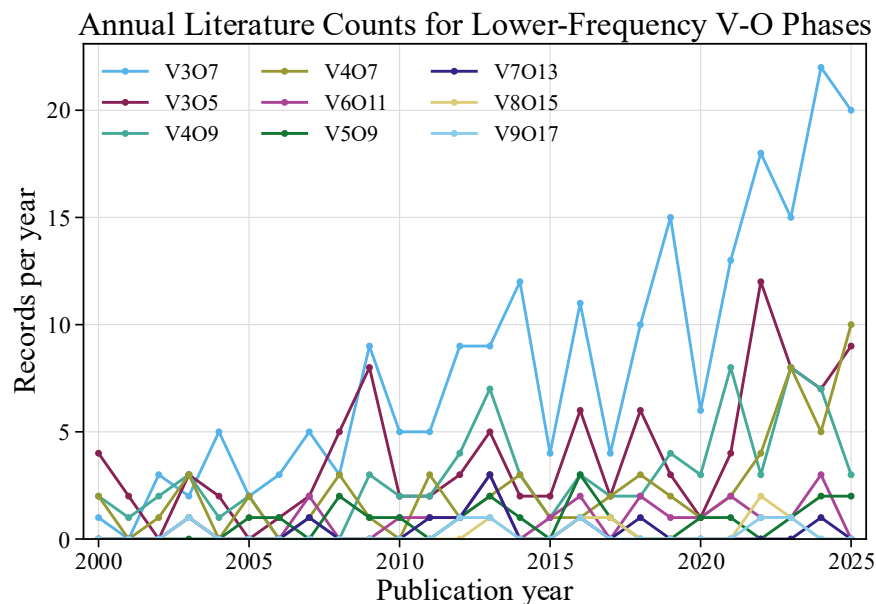


Figure S3: Annual publication counts (2000-2025) for lower-frequency V-O compositions, including Magneli-series phases. Although absolute counts remain modest, sustained growth in V3O7, V4O9, V6O11, and related stoichiometries indicates increasing research interest in less-explored phases of the V-O system

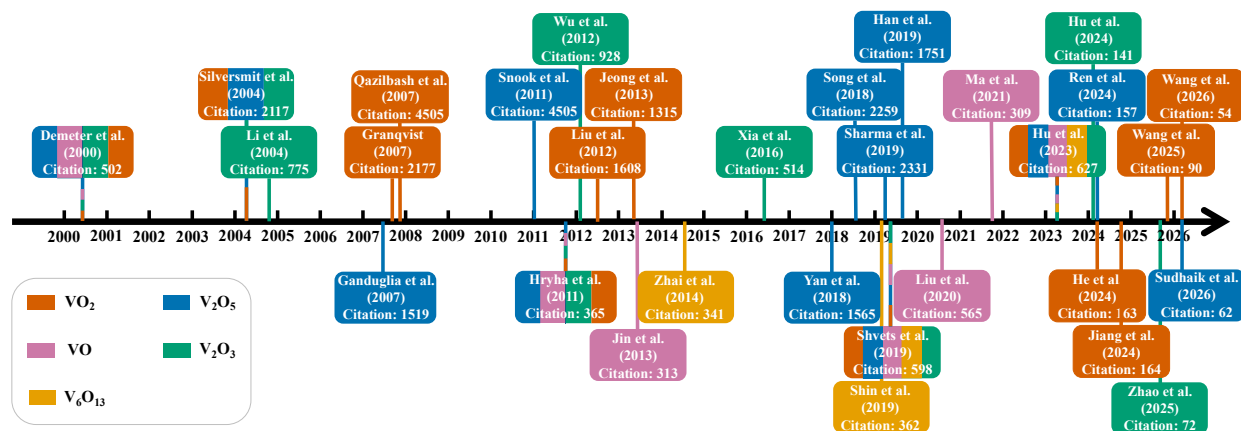


Figure S4: Timeline of the most highly cited publications for the five most-studied vanadium–oxide compositions (V_2O_5 , VO_2 , V_2O_3 , VO , and V_6O_{13}), positioned by publication year over 2000–2026. Each callout reports the first author, year, and Web of Science citation count; box colors correspond to the composition palette used in Fig. S2

S2. Conceptual overview of the inverse-design workflow

To complement the detailed pipeline of Fig. 3 in the main text, Figure S5 presents a conceptual abstraction of the inverse-design workflow, organized into its four functional stages: data, representation, coding, and training. This view emphasizes the logical flow of information, from the curated and substitution-augmented V-O database, through the element-resolved voxel representation, into the variational autoencoder that learns a compact latent space, and finally to the Wasserstein adversarial sampler whose decoded outputs are passed to first-principles validation.

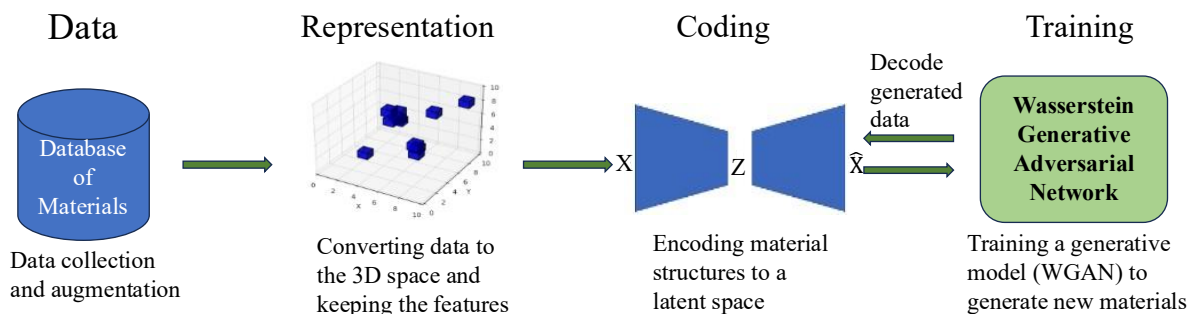


Figure S5: Conceptual schematic of the inverse-design workflow used in this study. Data: a curated database of V-O crystal structures is assembled from the Materials Project and augmented by element substitution. Representation: each structure is encoded as a set of element-resolved three-dimensional voxel grids that preserve geometric and chemical features in a form compatible with convolutional architectures. Coding: a variational autoencoder compresses the voxel grids into a low-dimensional latent space z and reconstructs them via the decoder. Training: a Wasserstein Generative Adversarial Network is trained to sample novel points from the learned latent space; sampled codes are decoded into candidate voxel grids and subsequently reconstructed into crystal structures for DFT validation

S3. Penalty-weight ablation

The formation-energy penalty weight λ in Eq. (4) governs the balance between distributional fidelity and the thermodynamic bias toward low-energy structures. Because full DFT validation of every generated set across all candidate λ values is computationally prohibitive, we used the frozen formation-energy regressor (test MAE 0.064 eV/atom) as a relative surrogate, applying the same stability criterion used for DFT screening to the predicted formation energies. Table S1 reports the resulting generation-quality metrics and predicted stability for $\lambda \in \{0, 0.05, 0.10, 0.20, 0.50\}$. Without the penalty ($\lambda = 0$), only 5.1% of structures are predicted stable; introducing the penalty raises this to 24.2% at $\lambda = 0.1$ and saturates thereafter (24.2–24.8% for $\lambda \geq 0.1$), while uniqueness, novelty, and structural diversity decrease monotonically with increasing λ and validity peaks at $\lambda = 0.1$. The production value $\lambda = 0.1$ therefore sits at the elbow of the stability-diversity trade-off, capturing essentially the full stability benefit at maximal validity and high diversity.

Table S1: Sensitivity of generation quality and predicted thermodynamic stability to the formation-energy penalty weight λ in Eq. (4). Predicted stability is the fraction of generated structures satisfying the stability criterion ($E_f < 0$ and $0 \leq \Delta E_{\text{hull}} \leq 300$ meV/atom above the MP convex hull) when E_f is estimated by the frozen CNN regressor (test MAE 0.064 eV/atom)

λ	Validity (%)	Uniqueness (%)	Novelty (%)	XRD diversity	Pred. stable (%)
0.00	76.7	97.5	99.6	0.983	5.1
0.05	80.6	97.3	99.6	0.981	12.4
0.10	81.5	96.6	99.5	0.973	24.2
0.20	79.7	95.5	99.0	0.960	24.8
0.50	73.2	89.5	98.7	0.889	23.7

S4. Vibrational and thermodynamic characterization of a sub-hull candidate

Figures S6 and S7 support the third phonon case study of Section 3.8 of the paper. Figure S6 shows the total phonon density of states of the below-hull V_2O_2 candidate. The spectrum extends from 0 to ~ 14 THz with no spectral weight at negative frequency, confirming the absence of soft modes anywhere in the Brillouin zone and complementing the band-path sampling of Fig. 10(c). Figure S7 shows harmonic thermodynamic functions obtained from the same force constants: Helmholtz free energy, entropy, and constant-volume heat capacity over 0 – 1000K. The heat capacity approaches the Dulong-Petit limit $3nR = 99.8 \text{ J K}^{-1} \text{ mol}^{-1}$ for the four-atom cell, and the free energy is finite and monotonically decreasing, both of which are only possible for a spectrum free of imaginary modes.

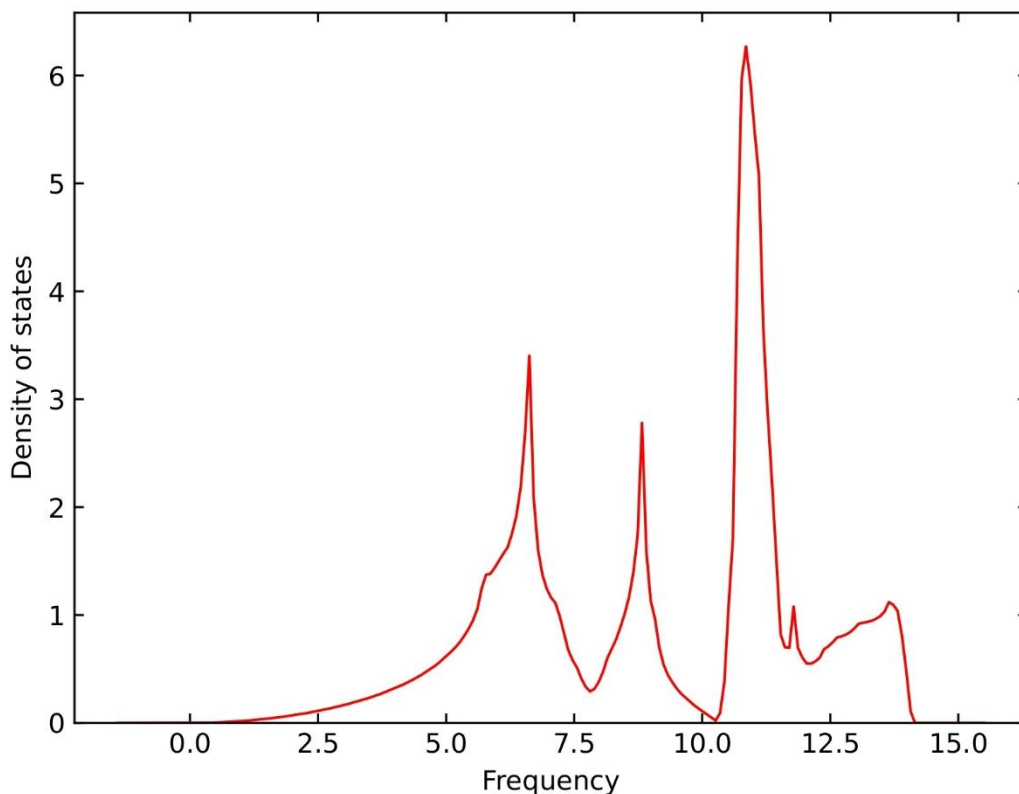


Figure S6: Total phonon density of states of the below-hull V_2O_2 candidate of Fig. 10(c). The spectrum spans 0 to ~ 14 THz with no spectral weight at negative frequency, confirming the absence of soft modes throughout the Brillouin zone and complementing the band-path sampling of Fig. 10(c) in the main text

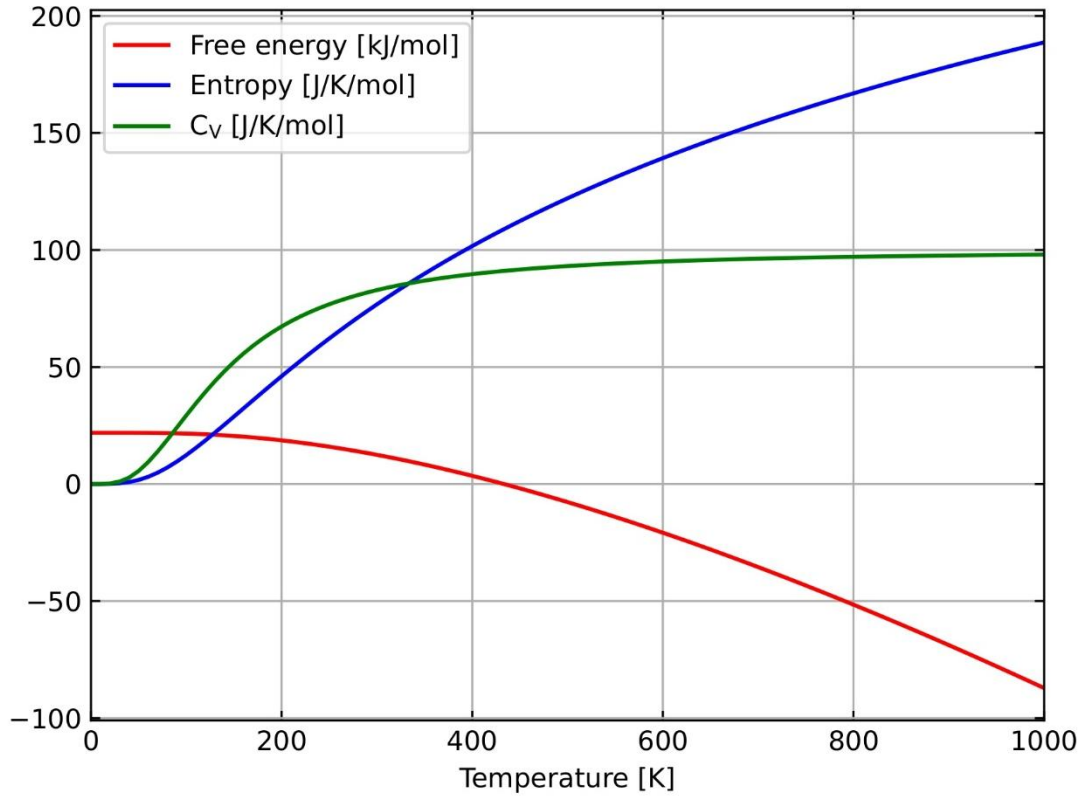


Figure S7: Harmonic thermodynamic functions of the below-hull V_2O_2 candidate, obtained from the same force constants as Figs. 10(c) and S6: Helmholtz free energy, entropy, and constant-volume heat capacity C_V over 0-1000K. The heat capacity saturates near $100 \text{ J K}^{-1} \text{ mol}^{-1}$, in agreement with the Dulong--Petit limit

Table S2: Phonon case studies spanning the three thermodynamic classes of Fig. 8. ΔE_{hull} is the distance to the MP-referenced convex hull; ω_{min} and ω_{max} are the minimum and maximum eigenfrequencies over the sampled band path

Composition	Class	Supercell	ω_{min} (THz)	ω_{max} (THz)	Verdict
V_2O_5	Unstable	$3 \times 3 \times 2$	≈ -20	≈ 140	Dynamically unstable
VO_2	Stable	$5 \times 5 \times 4$	≈ -2	≈ 23	Stable; shallow soft modes
V_2O_2	Below hull	$4 \times 4 \times 4$	≈ 0	≈ 13.5	Dynamically stable