# A UNIFIED SVD PERSPECTIVE: DECONSTRUCTING, EVALUATING, AND IMPROVING MODEL MERGING WITH ORTHO-MERGE

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Model merging is a powerful training-free technique for integrating the capabilities of multiple fine-tuned models, yet prevailing approaches—parameter-statistical (e.g., Average, TIES, DARE) and spectral/SVD-based (e.g., iso_c, KnOTS)—arise from disparate philosophies without a unifying account. We present a unified SVD-centric framework grounded in four principles—energy preservation, cross-task interference, spectral entropy, and information loss—that provides a consistent lens for analyzing merging algorithms. Guided by this framework, we introduce ORTHO-MERGE, a sign-aware deconflict-then-harmonize method. For each layer and task vector, we perform SVD and use signed similarities between leading singular directions to detect both redundant ($> \tau$) and oppositional ($< -\tau$) interference across tasks. The weaker singular component in each interfering pair is removed from its source task vector; the deconflicted vectors are then aggregated and harmonized via iso_c-style spectral averaging (SVD with mean-singular-value equalization). This training- and data-free pipeline resolves geometric conflicts before aggregation and controls the merged spectrum, preserving informative mid-rank structure while avoiding over-flattening. Across three CLIP backbones (ViT-B/32, ViT-B/16, ViT-L/14) and task suites of size 8/14/20, ORTHO-MERGE achieves state-of-the-art or competitive results on both average absolute and normalized accuracy. Spectrum diagnostics further show reduced spectral entropy and lower information loss, aligning the observed gains with our framework.

## 1 INTRODUCTION

Model merging is an increasingly pivotal technique that creates a single, powerful model by combining the weights of multiple, independently fine-tuned neural networks. This approach directly addresses a fundamental challenge in the current AI landscape: how to efficiently consolidate the specialized knowledge scattered across a vast ecosystem of expert models. The primary advantage of model merging lies in its remarkable efficiency. As a training-free process, it entirely bypasses the need for costly retraining, saving immense computational resources and time (Yang et al., 2024). Furthermore, its data-free nature makes it exceptionally valuable, and often the only viable option, in real-world scenarios where access to original training datasets is restricted due to privacy, proprietary, or logistical constraints (Ruan et al., 2025). By offering a pragmatic method to fuse diverse capabilities without relying on underlying data, model merging provides a scalable and robust solution for enhancing model performance and versatility across numerous applications.

Despite the empirical success of model merging, the field has evolved along two philosophically distinct trajectories. The first, operating from a statistical view, focuses on the numerical properties of model parameters. Methods such as simple weight averaging (Model Soup) (Wortsman et al., 2022), TIES-Merging (Yadav et al., 2023), DARE (Yu et al., 2024), and Breadcrumbs merging (Davari & Belilovsky, 2024) as a problem of mitigating destructive interference at the parameter level. They employ heuristics like pruning low-magnitude weights or resolving sign conflicts to harmonize conflicting parameter updates. The second trajectory adopts a geometric view, leveraging techniques like Singular Value Decomposition (SVD) in methods such as iso_c (Marczak et al., 2025)and KnOTS (Stoica et al., 2024). These approaches prioritize the preservation of structural geometry within the weight space, aiming to align the underlying structure of task vectors before fusion (Ruan et al., 2025). However, these two schools of thought have progressed largely in isolation. The field critically lacks a unified theoretical framework to connect them, causing progress to appear ad-hoc and preventing a systematic understanding of why certain methods succeed or fail. Consequently, significant performance bottlenecks remain, as existing techniques are unable to holistically resolve both parameter-level conflicts and structural-geometric incompatibilities.

To bridge this theoretical gap and address the aforementioned performance bottlenecks, we first propose a unified analytical framework for model merging. This framework is founded on four principles—energy preservation, cross-task interference, spectral entropy, and information loss—that provide a consistent lens for systematically evaluating the strengths and weaknesses of merging algorithms. By quantifying these aspects, our framework moves beyond ad-hoc heuristics and supplies principled guidance for designing more robust techniques.

Motivated by these insights, we introduce ORTHO-MERGE, a deconflict-then-harmonize algorithm that matches our implementation. For each layer and each task vector, we perform SVD and use signed similarities between leading singular directions to detect both redundant ($> \tau$) and oppositional ($< -\tau$) interference across tasks. Whenever an interfering pair is identified, the weaker singular component is removed from its source task vector, yielding a deconflicted set of task vectors. We then aggregate these deconflicted vectors and apply an iso_c-style spectral harmonization—SVD followed by mean–singular-value equalization—to control the merged spectrum without over-flattening. This deconflict before aggregate, then harmonize design preserves informative mid-rank structure and directly addresses the limitations of prior art that aggregate on conflict-laden inputs.

**Contributions**

- We present a unified SVD-centric framework that links the statistical and geometric views of model merging via four quantitative lenses: energy preservation, cross-task interference, spectral entropy, and information loss.
- Using this framework, we diagnose why existing methods degrade—showing that aggregation on misaligned task vectors induces both parameter-level interference and geometric conflicts.
- We propose ORTHO-MERGE, a deconflict-then-harmonize algorithm: per layer, we perform sign-aware SVD deconfliction by removing the weaker singular component whenever the signed similarity between leading directions exceeds a threshold, then aggregate the deconflicted vectors and apply iso_c-style spectral harmonization (SVD with mean–singular-value equalization).
- Extensive experiments across CLIP backbones (ViT-B/32, ViT-B/16, ViT-L/14) and 8/14/20-task suites show state-of-the-art or competitive results: ORTHO-MERGE improves average absolute and normalized accuracy while reducing spectral entropy and information loss.

## 2 RELATED WORK

A prominent and direct class of model merging techniques operates on the core assumption that the magnitude of a parameter change correlates with its importance to a specific task. These methods aim to mitigate interference by pruning or resetting less significant parameters before aggregation. Among the simplest of these is random pruning, exemplified by DARE (Drop And Rescale) (Yu et al., 2024). The core idea of DARE is to randomly drop a certain proportion of the delta parameters—the differences between fine-tuned and pre-trained weights—and then rescale the remaining ones to preserve their original collective magnitude. While this approach is straightforward and effective at moderate pruning rates, its performance can degrade significantly under extreme sparsity conditions. More deterministic, threshold-based pruning methods have also been proposed. TIES-Merging, for instance, first prunes parameters with low magnitudes and subsequently resolves sign conflicts among the remaining parameters across different models. Similarly, Breadcrumbs (Davari & Belilovsky, 2024) employs a dual masking strategy that simultaneously removes both small, noisy perturbations and large outliers from the task vectors, offering a more balanced approach to parameter selection.

Despite their effectiveness in reducing certain types of parameter conflicts, these magnitude-based methods share a fundamental limitation. They operate from an inherently statistical perspective, focusing on the numerical values of individual parameters while largely ignoring the deeper geometric and structural relationships within the task vectors. This oversight can lead to the disruption of the model's intrinsic functional structure during the merging process, creating a performance ceiling that motivates the exploration of alternative, structure-aware approaches.

In contrast to methods that focus on parameter magnitudes, a second approach model merging from a geometric perspective (Gargiulo et al., 2025). These techniques leverage spectral methods, such as Singular Value Decomposition (SVD), to analyze and combine models. The central idea is that the functional behavior of a model is better captured by the geometric structure and orientation of its weight matrices, rather than by the raw values of its parameters. A representative example of this approach is Isotropic Merging (iso_c) (Marczak et al., 2025). Its core mechanism involves first aggregating multiple task vectors, typically by summing them into a single representative vector. SVD is then performed on this combined vector, and its singular values are subsequently averaged or equalized. This process can be understood as a form of "spectral harmonization," where the principal geometric directions are preserved while the associated energy is uniformly distributed across them. Another prominent direction is seen in methods like KnOTS (Stoica et al., 2024), which is specifically designed for merging LoRA modules. KnOTS first maps the weight updates from multiple LoRA modules into an aligned, common subspace. Within this shared space, standard merging strategies like averaging or Task Arithmetic can then be applied more effectively. The development of such methods underscores the growing importance of structure-aware merging.

While these SVD-based methods are more adept at preserving the geometric integrity of the models, they possess a critical underlying limitation. They typically perform aggregation on task vectors that may already be in conflict. This means that if two task vectors are geometrically oppositional or incompatible, a simple averaging or harmonization process may neutralize or severely weaken the important signals from both tasks. This failure to resolve conflicts before aggregation creates a clear need for a more sophisticated approach that can deconflict task-specific knowledge prior to merging.

In summary, training-free merging has evolved along two tracks: magnitude-based (statistical) and spectral (geometric). What has been missing is a unified account that addresses both parameter-level interference and geometric incompatibility within one procedure. Our SVD-based framework provides this link and motivates ORTHO-MERGE: a mediate-then-harmonize pipeline that matches our implementation. Instead of pruning or discarding singular directions, ORTHO-MERGE performs a lightweight geometric mediation that optimizes small corrections to task vectors to reduce pairwise inner-product conflicts before aggregation. We then aggregate the mediated vectors and apply iso-C style spectral harmonization (SVD with mean singular-value equalization). This optimization-based deconfliction preserves informative mid-rank structure and creates favorable conditions for the subsequent spectral step, systematically overcoming the limitations observed in prior work.

# 3 Deconstructing Model Merging: A Unified SVD Framework

## 3.1 The Central Challenge of Model Merging: Cross-Task Conflict

A central challenge underpinning all model merging endeavors is the resolution of cross-task conflicts. These conflicts arise because independently fine-tuned models, starting from the same pre-trained initialization, learn distinct sets of parameter adjustments to solve their respective tasks. When naively combined, these adjustments can interfere with or destructively cancel each other out, leading to a merged model with degraded performance. This phenomenon manifests in both the parameter space and the geometric space of the model weights.

From a statistical viewpoint, conflict is observable at the individual parameter level. For any given weight in a neural network, the corresponding changes encoded in two different task vectors, $\Delta\mathbf{\Theta}_A$ and $\Delta\mathbf{\Theta}_B$, may exhibit sign opposition (e.g., one positive and the other negative) or significant magnitude discrepancy.

The detrimental effect of such conflicts is most evident in simple averaging. Define the averaged update

$$\Delta\mathbf{\Theta}_{\text{avg}} = \frac{\Delta\mathbf{\Theta}_A + \Delta\mathbf{\Theta}_B}{2}, \tag{1}$$

whose squared Frobenius norm (total energy) is

$$\left\|\Delta\mathbf{\Theta}_{\text{avg}}\right\|_F^2 \;=\; \tfrac{1}{4}\Big(\left\|\Delta\mathbf{\Theta}_A\right\|_F^2 + \left\|\Delta\mathbf{\Theta}_B\right\|_F^2 + 2\ \mathrm{Tr}\big(\Delta\mathbf{\Theta}_A^\top \Delta\mathbf{\Theta}_B\big)\Big). \tag{2}$$

The term $2\ \mathrm{Tr}\big(\Delta\mathbf{\Theta}_A^\top \Delta\mathbf{\Theta}_B\big)$ captures the interference between the two updates. When sign conflicts are prevalent, the inner product becomes negative, inducing destructive interference that reduces the total energy and thus erodes learned information. This observation is consistent with loss-landscape analyses: independently trained networks often lie in different re-basins(Rinaldi et al., 2025), and linear interpolation between their parameters (as in averaging) can cross a high-loss region, breaking Linear Mode Connectivity (LMC)(Ito et al., 2025) and yielding suboptimal merged performance.

Statistical methods attempt to mitigate conflicts by manipulating parameter values under the assumption that statistics like magnitude are reliable proxies for importance. However, this approach is inherently "blind" to the underlying geometric structure of the task vectors. The core ideas of prominent methods include:

- **TIES-Merging**(Yadav et al., 2023) attempts to denoise task vectors by pruning low-magnitude weights and then mitigates direct opposition by resolving sign conflicts among the remaining parameters.
- **DARE (Drop And Rescale)** (Yu et al., 2024)directly addresses the energy loss from sparsification. It randomly drops a fraction of parameters and then applies a compensatory rescaling to the remainder to preserve the vector's overall magnitude.
- **Breadcrumbs**(Davari & Belilovsky, 2024) employs a more sophisticated heuristic by removing both the highest-magnitude outliers and the lowest-magnitude noise. This can be interpreted as an attempt to flatten the spectral distribution of the weights to improve generalization.

Despite their increasing sophistication, these methods all operate indirectly on the model's geometry, lacking the surgical precision to control the structural components directly.

From a geometric perspective, where task vectors are analyzed via Singular Value Decomposition (SVD), conflict manifests as a lack of orthogonality between the singular vector subspaces of the task vectors. Each task vector can be decomposed into its constituent geometric components:

$$\Delta\mathbf{\Theta}_A = U_A \mathbf{\Sigma}_A V_A^\top, \qquad \Delta\mathbf{\Theta}_B = U_B \mathbf{\Sigma}_B V_B^\top. \tag{3}$$

Many merging methods operate on the sum of task vectors,

$$\Delta\mathbf{\Theta}_{\text{sum}} \;=\; \Delta\mathbf{\Theta}_A + \Delta\mathbf{\Theta}_B. \tag{4}$$

As shown in the block matrix formulation,

$$\Delta\mathbf{\Theta}_{\text{sum}} \;=\; [U_A \quad U_B] \begin{bmatrix} \mathbf{\Sigma}_A & 0 \\ 0 & \mathbf{\Sigma}_B \end{bmatrix} \begin{bmatrix} V_A^\top \\ V_B^\top \end{bmatrix}. \tag{5}$$

Crucially, Equation 5 is not a valid SVD: a formal SVD requires the left and right factor matrices to be orthogonal, i.e., their columns/rows form orthonormal bases. However, the concatenated matrix $\big[\, U_A \;\; U_B \,\big]$ is orthogonal only when the column spaces of $U_A$ and $U_B$ are mutually orthogonal—an exceptional case. In practice, nonzero subspace similarity between tasks indicates this condition rarely holds. Therefore, summing task vectors before analysis amounts to operating on an entangled, non-orthogonal basis, which mixes and distorts task-specific geometric structure and can lead to loss of critical information.

To empirically demonstrate that cross-task conflict is a pervasive issue, we analyze the geometric similarity between the task vectors derived from a ViT-B/16 model fine-tuned on eight distinct datasets: SVHN, MNIST, DTD, Cars, EuroSAT, GTSRB, RESISC45, and SUN397. Our analysis focuses on a representative deep layer—the 10th MLP layer—and measures the alignment of its most informative subspaces, which we define as the spaces spanned by the top-$k$ left singular vectors of the weight delta matrices $\Delta\Theta$. Formally, for this layer, we take the task vectors from two tasks, $\Delta\Theta_A$ and $\Delta\Theta_B$, and perform SVD to obtain their respective top-$k$ left singular vector bases, $U_A$

and $U_B$. The similarity, avg_sim, is calculated as the average of the singular values of the overlap matrix $S = U_A^\top U_B$.

The resulting similarity matrix is visualized as a heatmap in Figure 1. The heatmap clearly shows that non-zero similarity exists between almost all task pairs, confirming their lack of orthogonality. Most notably, we observe a high degree of similarity between semantically related tasks, such as MNIST and SVHN (similarity = 0.37), both of which are digit recognition datasets. A significant, albeit smaller, similarity is also observed between SVHN and GTSRB (similarity = 0.22). This provides undeniable evidence that significant structural conflict and redundancy exist between task vectors. Any merging method that naively aggregates these vectors without explicitly addressing such high-similarity interference is likely to suffer from destructive interference and suboptimal performance.

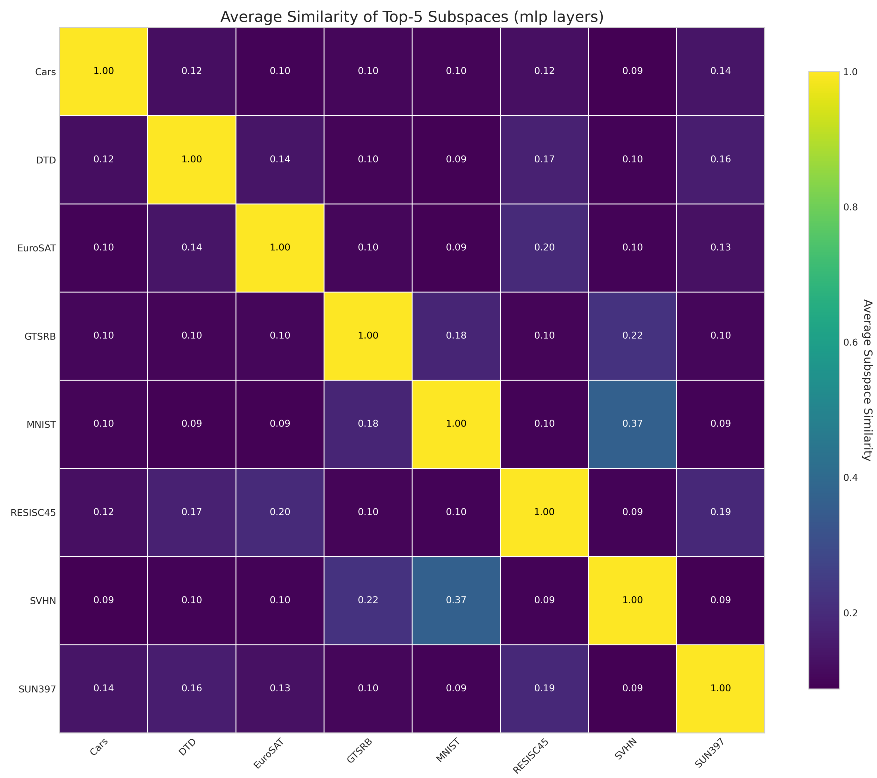


Figure 1: Top-5 subspace similarity across tasks for the ViT-B/16 block-10 MLP task vectors. For each task vector at this layer, we compute its SVD and use the leading five singular vectors to span a 5D representative subspace.

To investigate the effects of merging highly similar tasks, we can analyze the process from a vector geometry perspective. To empirically validate our theory, we designed a selective integration experiment. We first create a base model task vector, $\Delta\theta_{AC}$, by averaging the vectors for SVHN (Task A) and the dissimilar texture dataset DTD (Task C):

$$\Delta\Theta_{AC} = \frac{1}{2}(\Delta\Theta_A + \Delta\Theta_C) \tag{6}$$

Next, we decompose the vector for a similar task, MNIST (Task B), into a sum of its rank-1 singular components using SVD:

$$\Delta\Theta_B = U_B \Sigma_B V_B^T = \sum_{j=1}^{r} c_j, \quad \text{where } c_j = \sigma_j u_j v_j^T \tag{7}$$

Here, $c_j$ is the $j$-th singular component, composed of the $j$-th singular value $\sigma_j$ and its corresponding left and right singular vectors $u_j$ and $v_j$. The task vector at each incremental injection step $i$, $\Delta\Theta^{(i)}$, is then defined by cumulatively adding these components to the base vector:

$$\Delta\Theta^{(i)} = \Delta\Theta_{AC} + \sum_{j=1}^{i} c_j \tag{8}$$

At each step, the resulting model is evaluated on the validation sets for SVHN and DTD. The results, plotted in Figure 2, provide a clear visualization of this duality. As the components of MNIST are injected, the performance on the related task, SVHN (blue line), remains remarkably stable. In stark contrast, the performance on the dissimilar task, DTD (orange line), degrades catastrophically. This demonstrates that injecting a similar task vector reallocates the model's capacity, reinforcing shared capabilities (digit recognition) at the expense of unrelated ones (texture classification), effectively "drowning out" the features required for Task C.

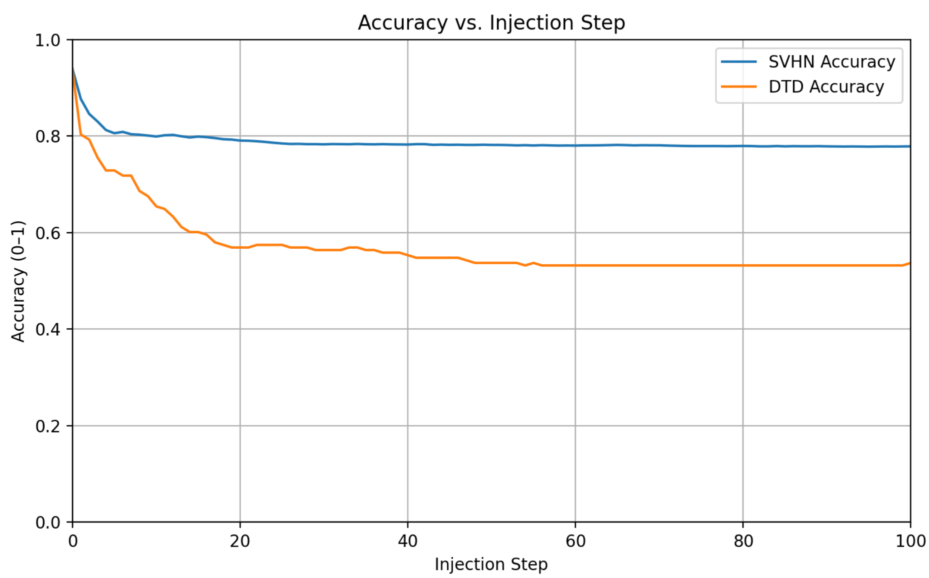


Figure 2: Accuracy vs. injection step on SVHN and DTD with components of MNIST injecting

### 3.2 An SVD-Based Unified Framework

Building on the established challenges of cross-task conflict, we now propose a unified framework that uses Singular Value Decomposition (SVD) as a "universal language" to analyze and interpret any merging operation. Any such operation can be fully characterized by its effect on a task vector's geometric structure $(U, V)$ and its energy spectrum $(\Sigma)$. A primary observation from the cumulative energy spectrum (Figure 3 ) is that task vectors exhibit a "low-rank head" structure, where a majority of their total energy is concentrated in the top few dozen singular values. While this confirms the importance of high-energy components, a deeper analysis reveals that the shape of the energy spectrum, particularly the distribution of energy in its long tail, is critical for the generalization capacity of the merged model.

To quantify this spectral flatness, we introduce the concept of spectral entropy. For a given spectrum $\Sigma$ with singular values $\{\sigma_i\}$, we define the normalized energy of the $i$-th component as $p_i = \sigma_i^2 / \sum_j \sigma_j^2$. The spectral entropy is then given by:

$$H(\Sigma) = -\sum_i p_i \log p_i \tag{9}$$

A low spectral entropy indicates a "peaky" spectrum where energy is concentrated in a few dominant components, leading to a narrow subspace that struggles to accommodate multiple tasks. Conversely, a high spectral entropy signifies a flatter, "long-tail" distribution, which we hypothesize correlates with better generalization. Figure 4 visualizes this effect, showing how naive averaging results in a collapsed, low-entropy spectrum, while more advanced methods produce spectra with higher entropy.

This SVD-based perspective allows us to "translate" and critique existing methods with greater precision. Statistical methods like DARE, expressed as $\Delta\Theta_{DARE} = s \cdot (\Delta\Theta \odot M_{mask})$, act as indirect spectral editors; their operations in the parameter space constitute a complex, non-linear transformation that affects all SVD components $(U, \Sigma, V)$ imprecisely. In contrast, methods like iso_c are direct spectral editors, precisely modifying only the energy spectrum $\Sigma$ of a pre-summed, conflict-ridden vector. This framework thus reveals a unified insight: statistical methods are imprecise manipulators of geometry and energy, while existing SVD methods are precise manipulators that operate on a flawed input. Therefore, an ideal method must first resolve the structural conflicts (between the $U$ and $V$ components of different tasks) before precisely harmonizing their energy spectra ($\Sigma$). This provides the core theoretical motivation for our proposed ORTHO-MERGE algorithm.

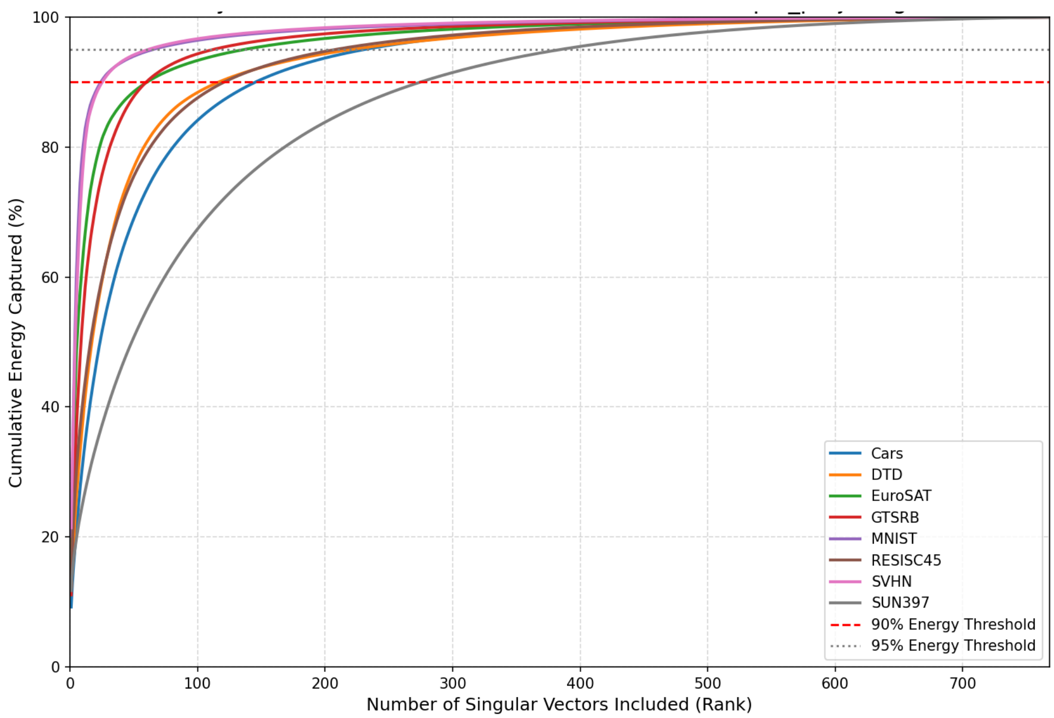


Figure 3: Cumulative energy captured by singular values versus rank for each task-specific model (ViT-B/16, block-10 MLP).

In model merging, different operational steps can lead to varying degrees of information loss. For instance, while a method like iso_c might excel in specific areas such as spectral flattening, the fundamental process of aggregating task vectors can inherently cause a loss of nuanced, task-specific information. While this loss can be measured in multiple ways, this work proposes using Kullback-Leibler (KL) Divergence to precisely evaluate the change at each step, which is critical for the rapid development of advanced merging techniques. This method defines information loss as the divergence between the output probability distributions of a sub-model before the merge and the new model after the merge.

Specifically, the information loss for a sub-model $i$ is quantified as the expected KL Divergence over a standard probe dataset, $D_{\text{probe}}$. The formula is given by:

$$\mathcal{L}_{\text{KL}}(i) = \mathbb{E}_{x \sim D_{\text{probe}}} \left[ D_{\text{KL}}(P_i(y|x) \,\|\, \overline{P}(y|x)) \right] \tag{10}$$

where the KL Divergence itself is calculated as:

$$D_{\text{KL}}(P_i \,\|\, \overline{P}) = \sum_{y \in \mathcal{Y}} P_i(y|x) \log \frac{P_i(y|x)}{\overline{P}(y|x)} \tag{11}$$

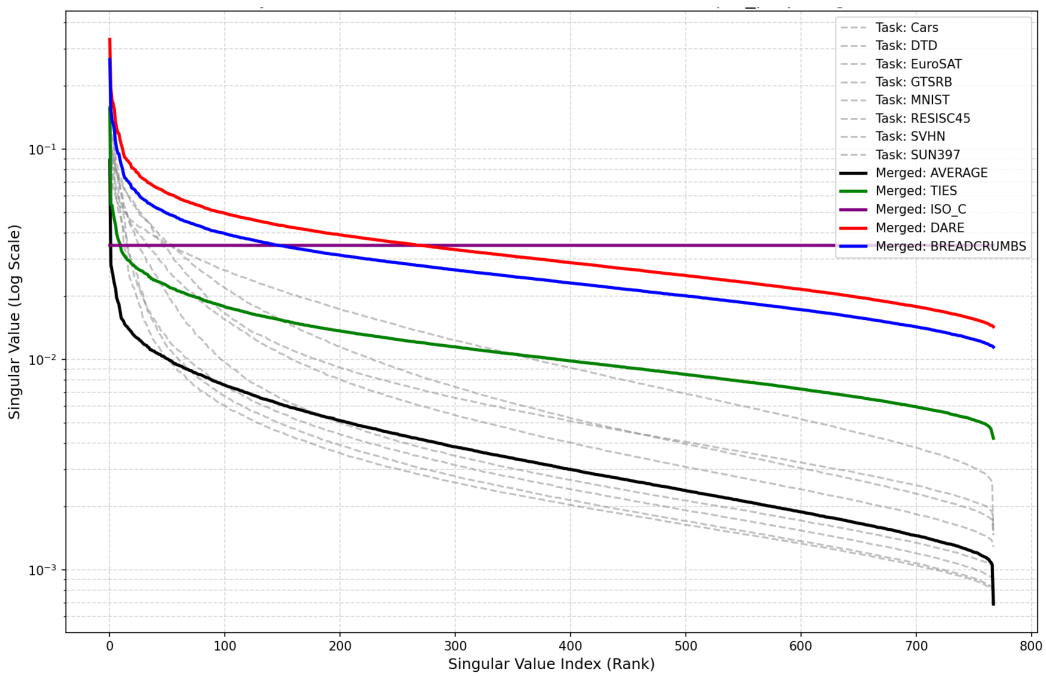


Figure 4: Log-scale singular-value spectra comparing merged methods on the same target layer (ViT-B/16, block-10 MLP).

Here, $P_i(y|x)$ is the original sub-model's probability distribution over the outputs $y \in \mathcal{Y}$ for a given input $x$, and $\overline{P}(y|x)$ is the distribution of the newly merged model. This metric asymmetrically measures the information lost when approximating the original distribution $P_i$ with the merged one $\overline{P}$, thus capturing subtle changes in model certainty and predictions. By applying this fine-grained analysis at every stage, we can better understand and mitigate information loss in complex model merging pipelines.

## 4 Method

For a given 2D weight layer, let the *task vectors* be $\{\Delta\Theta_t\}_{t=1}^{T}$, one per fine-tuned model. For each task $t$, we take the compact SVD

$$\Delta\Theta_t \;=\; U_t\,\Sigma_t\,V_t^{\top} \;=\; \sum_{s=1}^{r_t} \sigma_{t,s}\,u_{t,s}\,v_{t,s}^{\top} \;=\; \sum_{s=1}^{r_t} c_{t,s}, \tag{12}$$

where $r_t = \operatorname{rank}(\Delta\Theta_t)$, $\Sigma_t = \operatorname{diag}(\sigma_{t,1},\ldots,\sigma_{t,r_t})$, $u_{t,s}$ and $v_{t,s}$ are the $s$-th left/right singular vectors, and $c_{t,s} \overset{\text{def}}{=} \sigma_{t,s}\,u_{t,s}\,v_{t,s}^{\top}$ is the $s$-th *singular component* of task $t$. We use $\langle\cdot,\cdot\rangle$ for the Euclidean inner product and consider the top-$k$ left directions $U_t^{(k)} = \big[u_{t,1},\ldots,u_{t,k}\big]$.

To prevent destructive aggregation on misaligned inputs, we first deconflict task vectors *before* summation. For any two tasks $(i,j)$ and indices $(m,n) \le k$, define the *signed similarity*

$$s_{ij}^{mn} \;=\; \big\langle u_{i,m},\, u_{j,n}\big\rangle \in [-1,1]. \tag{13}$$

When $|s_{ij}^{mn}| > \tau$ (redundant if $> \tau$, oppositional if $< -\tau$), we remove the *weaker* of the two interfering components from its source task vector:

$$\text{if }\ \sigma_{i,m} < \sigma_{j,n} \;\Rightarrow\; \Delta\Theta_i \;\leftarrow\; \Delta\Theta_i - c_{i,m} \quad \text{else} \quad \Delta\Theta_j \;\leftarrow\; \Delta\Theta_j - c_{j,n}. \tag{14}$$

After scanning all task pairs and $(m,n) \le k$, we obtain a *deconflicted* set $\{\widehat{\Delta\Theta}_t\}_{t=1}^{T}$.

We then aggregate the deconflicted task vectors and apply a single spectral smoothing step:

$$\Delta\Theta_{\text{sum}} = \sum_{t=1}^{T} \widehat{\Delta\Theta}_t \;=\; U\,\Sigma\,V^{\top}, \tag{15}$$

$$\bar{\sigma} = \frac{1}{r}\sum_{q=1}^{r} \sigma_q, \qquad \Delta\Theta_{\text{new}} \;=\; U\,\mathrm{diag}(\bar{\sigma}\,\mathbf{1})\,V^{\top}, \tag{16}$$

where $r = \mathrm{rank}(\Delta\Theta_{\text{sum}})$. Eq. equation 16 corresponds to iso_c-style harmonization (SVD followed by mean–singular-value equalization), which controls the spectrum without over-flattening. Non-2D tensors (e.g., biases, embeddings, text_projection) are averaged.

## 5 Experiments and results

We evaluate our approach on three CLIP visual backbones (ViT-B/32, ViT-B/16, ViT-L/14) and three task suites with cardinality 8, 14, and 20, following the training-free merging protocol of prior work. Baselines include Weight Averaging, Task Arithmetic, TIES/Consensus TA, TSV-M, and recent spectral methods (iso_c/iso_cts); for reference we also report the zero-shot model (lower bound) and the average of individually fine-tuned models (upper bound). Performance is summarized by average absolute accuracy and average normalized accuracy. Across all backbones and task suites, our method delivers the best or tied-best results.

Table 1: Average absolute accuracy (%) and normalized accuracy (%) for 8/14/20-task suites across three backbones.

| Backbone | ViT-B/32 | | | | | ViT-B/16 | | | | | | ViT-L/14 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suite | 8 tasks | | 14 tasks | 20 tasks | | 8 tasks | | 14 tasks | | 20 tasks | | 8 tasks | | 14 tasks | | 20 tasks | |
| Metric | Abs | Norm | Norm | Abs | Norm | Abs | Norm | Abs | Norm | Abs | Norm | Abs | Norm | Abs | Norm | Abs | Norm |
| Zeroshot | 46.2 | 51.3 | 60.2 | 53.1 | 58.4 | 52.4 | 57.2 | 58.8 | 63.9 | 56.3 | 61.5 | 61.6 | 65 | 65.2 | 70.5 | 64.3 | 67.8 |
| Weight Averaging | 63.4 | 69.3 | 68.2 | 59.4 | 62.4 | 66.6 | 70.6 | 67.7 | 71.2 | 63.6 | 68.3 | 77.5 | 80.1 | 74.5 | 79.1 | 67.8 | 71.5 |
| Task Arithmetic | 67.9 | 71.5 | 68.9 | 60.5 | 66.7 | 72.4 | 76.5 | 68.5 | 72.8 | 62.7 | 67.6 | 80.3 | 85.6 | 77.4 | 80.5 | 71.6 | 75 |
| Ties | 72.3 | 76.8 | 74.6 | 62.4 | 68.2 | 72.6 | 76.5 | 69.3 | 74.1 | 65.2 | 70.1 | 83.4 | 87.1 | 80.1 | 83.6 | 77.4 | 80.3 |
| TSV-M | 82.6 | 87.2 | 82.1 | 74.5 | 80.9 | 77 | 83.4 | 80.4 | 86.5 | 77.3 | 83.2 | 89.2 | 92.2 | 85.3 | 90.1 | 84.5 | 88.6 |
| iso_c | 82.8 | 87.6 | 84.5 | 79.9 | 83.1 | 81.6 | 85.3 | 81.3 | 86.4 | 77.6 | 83.3 | 89.1 | 92 | 85.8 | 90.9 | 85.2 | 89.6 |
| iso_cts | 83.5 | 88.6 | 85.8 | 80.5 | 84.8 | 83 | 87.1 | 81.4 | 86.6 | 79.5 | 85.2 | 89.2 | 92.5 | 86.1 | 91.2 | 86.3 | 90.4 |
| ORTHO-MERGE(our) | 83.1 | 87.6 | 85.4 | **80.5** | **84.9** | **83.1** | **87.2** | 81.3 | 86.5 | **79.6** | **85.3** | **89.4** | **92.7** | **86.1** | **91.2** | **86.6** | **90.6** |

## 6 Conclusion

We presented a unified SVD view of training-free model merging and proposed ORTHO-MERGE, a sign-aware deconflict-then-harmonize pipeline. Per layer, we perform SVD on each task vector, identify redundant/oppositional interference via signed similarities of leading directions, remove the weaker singular component, and then aggregate and apply iso_c-style spectral harmonization. Spectral diagnostics (spectral entropy, information loss) indicate that ORTHO-MERGE preserves informative mid-rank structure while avoiding over-flattening. Across three CLIP backbones (ViT-B/32, ViT-B/16, ViT-L/14) and 8/14/20-task suites, it achieves state-of-the-art or competitive accuracy under both absolute and normalized metrics, narrowing the gap to individually fine-tuned models without training or data. Future work includes adaptive, data-free criteria for interference detection (e.g., learned or layer-wise thresholds), soft reweighting in place of hard removals, and extending analysis to right-singular alignments and non-2D tensors.

## References


MohammadReza Davari and Eugene Belilovsky. Model breadcrumbs: Scaling multi-task model merging with sparse masks. In *European Conference on Computer Vision*, pp. 270–287. Springer, 2024.

Antonio Andrea Gargiulo, Donato Crisostomi, Maria Sofia Bucarelli, Simone Scardapane, Fabrizio Silvestri, and Emanuele Rodola. Task singular vectors: Reducing task interference in model merging. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 18695–18705, 2025.

Akira Ito, Masanori Yamada, and Atsutoshi Kumagai. Linear mode connectivity between multiple models modulo permutation symmetries. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=qaJuLzY6iL.

Daniel Marczak, Simone Magistri, Sebastian Cygert, Bartłomiej Twardowski, Andrew D Bagdanov, and Joost van de Weijer. No task left behind: Isotropic model merging with common and task-specific subspaces. *arXiv preprint arXiv:2502.04959*, 2025.

Filippo Rinaldi, Giacomo Capitani, Lorenzo Bonicelli, Donato Crisostomi, Federico Bolelli, Elisa Ficarra, Emanuele Rodola, Simone Calderara, and Angelo Porrello. Update your transformer to the latest release: Re-basin of task vectors. *arXiv preprint arXiv:2505.22697*, 2025.

Wei Ruan, Tianze Yang, Yifan Zhou, Tianming Liu, and Jin Lu. From task-specific models to unified systems: A review of model merging approaches. *arXiv preprint arXiv:2503.08998*, 2025.

George Stoica, Pratik Ramesh, Boglarka Ecsedi, Leshem Choshen, and Judy Hoffman. Model merging with svd to tie the knots. *arXiv preprint arXiv:2410.19735*, 2024.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pp. 23965–23998. PMLR, 2022.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in Neural Information Processing Systems*, 36:7093–7115, 2023.

Enneng Yang, Li Shen, Guibing Guo, Xingwei Wang, Xiaochun Cao, Jie Zhang, and Dacheng Tao. Model merging in llms, mllms, and beyond: Methods, theories, applications and opportunities. *arXiv preprint arXiv:2408.07666*, 2024.

Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *Forty-first International Conference on Machine Learning*, 2024.

## A Appendix

You may include other additional sections here.