Similarity Across Layers and Resolutions: A Theory of Difference, Identity, and Asymmetric Transport

Yuval Fradkin

26 August 2026

Abstract

We develop a mathematical framework in which *similarity*, *difference*, and *identity* are resolution-dependent and layer-dependent relations rather than absolute properties of object pairs. The central object is $S(A, B; \ell, r)$, where ℓ is a descriptive layer and r is observational resolution within that layer. A fundamental distinction is established between *intrinsic* and *observable* difference: reducing resolution changes which differences can be resolved, not the differences that exist. This yields the decomposition $D_{\text{total}} = D_{\text{vis}}(r) + D_{\text{hid}}(r)$, formalized in a Hilbert space spectral setting. The layer space $\mathcal{L}$ is given a formal structure via six axioms, enabling the main new result: a *Layer–Resolution Monotonicity Theorem* (Theorem 7.3) which bounds how similarity can vary when resolution and layer change simultaneously. The *Layered Similarity Profile* $\mathbf{S}_{A,B}(r) = \{S(A, B; \ell, r) : \ell \in \mathcal{L}\}$ is introduced as the primary quantitative object; scalar aggregation is treated as secondary and measure-dependent. Identity is defined as the limit case of complete similarity: $A \equiv_{\ell,r} B \iff S(A, B; \ell, r) = 1$. Two new results complete the cross-layer transport theory: the *Downward Transfer Theorem* (Theorem 11.1) gives $S_{\ell_1} \geq S_{\ell_2}^{\kappa}$ (exponential kernel); Proposition 11.6 proves global upward transfer is impossible when $\ker \pi \neq \{0\}$; and Theorem 11.9 (Conditional Upward Transfer) shows that bounding the resolution-visible kernel component $\|P_{\ker \pi} R_r^{(\ell_2)} u_2\| \leq \varepsilon$ recovers $S_{\ell_2} \geq \Phi(D_{\ell_1}/\sigma + \varepsilon)$, where $\sigma = \sigma(\pi_{\ell_1,\ell_2})$ is the minimum gain of the coarsening map on $(\ker \pi)^{\perp}$. Together these form an intrinsically asymmetric transport theory. Two further results extend the theory to approximate settings: Theorem 11.11 (Robust Downward Transport) replaces exact axiom (L6) with an approximate intertwining error δ; Corollary 11.13 (Relative Robust Transport) measures the error relative to the post-resolution norm, yielding $S_{\ell_1} \geq S_{\ell_2}^{\kappa+\delta_r}$ — a robust similarity-to-similarity law that degrades continuously to the exact theory as $\delta_r \to 0$. Supporting structures – resolution operators, partition lattice, ultrametric geometry, statistical resolution, and resource-constrained distinguishability – are developed as derived realizations of the

four-level hierarchy: **Layer** → **Resolution** → **Accessible Difference** → **Similarity / Identity**.

Preliminary research paper (version 45). All mathematical claims are stated with explicit hypotheses. Version 45: (1) rephrases the admissibility clause in Step 3 of Theorem 11.21 to make $Ru = v \neq 0$ unambiguous; (2) sharpens the necessity direction of Proposition 8.2 to rest directly on the requirement that the denominator $\nu(\mathcal{L})$ be a finite real number, removing the $S \equiv 1$ witness.

Contents

1. Introduction

Conventional treatments of similarity take one of two forms. A *geometric* approach defines similarity via a distance in some fixed feature space; an *axiomatic* approach (following Tversky) imposes constraints on a scalar function $S(A, B)$ without specifying its computational form. Both approaches share an implicit assumption: that there exists a canonical description at which comparison should be performed, and that the result is a single number.

This paper argues that both assumptions are too strong.

First, the level at which objects are described – atomic, molecular, cellular, anatomical, physiological, behavioral, cognitive, social – is an independent parameter, not a fixed background. Two genetically identical twins are indistinguishable at the genomic layer; they may be entirely distinct at the behavioral layer. Treating description level as a parameter, not a constant, is the purpose of the *layer* coordinate ℓ.

Second, even within a fixed layer, the amount of distinguishing information available to an observer varies. An image compressed to ten pixels cannot reveal distinctions that are visible at full resolution. A measurement with high resource cost may resolve differences that a cheap measurement cannot. This is the purpose of the *resolution* coordinate r.

Example 1.1 (The Human–Cat Observation)**.** A human and a domestic cat are clearly distinguishable at standard biological resolution. At very low resolution – large physical distance, extreme image compression, or maximally coarse-grained description – many features that separate them cease to be accessible. The objects do not become more similar; rather, fewer of their differences remain *observable*. The total difference is conserved; only its visible fraction shrinks.

The key conceptual move is stated precisely as a principle in Section 2 and formalized mathematically in Section 5.

Contributions of this paper.

(i) A six-axiom formal structure for the layer space $\mathcal{L}$ (Section 3): axioms (L1)–(L4) provide the order-theoretic and separability conditions; axiom (L5) – Contractive Layer Coarsening – supplies the metric condition for Step 1 of Theorem 7.3; axiom (L6) – Layer–Resolution Compatibility – supplies the typed intertwining condition for Step 2.

(ii) A new theorem – the *Layer–Resolution Monotonicity Theorem* (Theorem 7.3) – that simultaneously involves the layer coordinate, the resolution coordinate, and the similarity function, and is not a consequence of any single one of those components.

(iii) A precise measure-theoretic treatment of the aggregation problem: when and how the Layered Similarity Profile can be legitimately collapsed to a scalar (Section 8).

(iv) The *Cross-Layer Transfer Theorem* (Theorem 11.1, Section 11.4): the first result bounding similarity at one layer from similarity at a related layer, via the layer contraction ratio $\kappa(\ell_1, \ell_2)$.

(v) Definitions and supporting results for the full hierarchy of constructs listed in the Appendix, with explicit statement of conditions wherever a claim depends on hypotheses.

What this paper does not claim. The partition lattice, ultrametric geometry, and spectral decompositions used here belong to established mathematics. The claim is not that these structures are novel, but that their *unified deployment around* $S(A, B; \ell, r)$ produces a framework that neither information theory, nor similarity theory, nor multilayer network theory currently supplies in the form developed here.

Organization. Section 2 states the central conceptual principle and the four-level hierarchy. Section 3 axiomatizes the layer space. Section 4 develops resolution operators, partitions, and lattices. Section 5 establishes the visible/hidden decomposition. Section 6 introduces the distinction space and refinement invariance. Section 7 develops multi-layer similarity, the Layered Similarity Profile, and the main theorem. Section 8 treats scalar aggregation. Section 9 defines identity as complete similarity. Section 10 collects supporting structures. Section 11 discusses open problems and research directions, including Section 11.4 which develops the Cross-Layer Transfer Theory and Section 11.5 which proves the Robust Downward Transport Theorem for approximate compatibility. Section 12 proposes a research direction for analogy discovery in mathematics. Section 13 describes a proposed application to object recognition in computer vision. Section C proposes a Layer–Resolution Similarity Controller for large language models. Section 11.7 provides a structured comparison with the closest existing frameworks.

2. The Central Framework

2.1. The Four-Level Hierarchy

The framework is organized around four nested conceptual levels:

$$\boxed{\textbf{Layer} \longrightarrow \textbf{Resolution} \longrightarrow \textbf{Accessible Difference} \longrightarrow \textbf{Similarity / Identity}.} \tag{1}$$

Every construct in the paper is positioned within this hierarchy. No construct that belongs to level k is treated as foundational at level $k-1$.

2.2. The Core Similarity Object

Definition 2.1 (Layer- and Resolution-Dependent Similarity)**.** Let X be a set of objects, $\mathcal{L}$ a layer space (axiomatized in Section 3), and $\mathcal{R}$ a resolution space (formalized in Section 4). A *layer- and resolution-dependent similarity function* is a map

$$S : X \times X \times \mathcal{L} \times \mathcal{R} \longrightarrow [0, 1]$$

satisfying:

(S1) *Self-similarity:* $S(A, A; \ell, r) = 1$ for all A, ℓ, r.

(S2) *Symmetry:* $S(A, B; \ell, r) = S(B, A; \ell, r)$ for all A, B, ℓ, r.

(S3) *Resolution monotonicity* (conditional): For any ℓ, if $r_1 \preceq r_2$ (meaning r_1 is coarser than r_2) then $S(A, B; \ell, r_1) \geq S(A, B; \ell, r_2)$, provided the resolution operators are nested orthogonal projections in the sense of Section 4.

Axiom (S3) is conditional; its hypotheses must be verified for each specific model.

Remark 2.2. The conventional function $S(A, B)$ appears as the special case where $\mathcal{L}$ is a singleton and $\mathcal{R}$ has a single element. The function $S(A, B; r)$ appears as the further special case where $|\mathcal{L}| = 1$. Neither special case is adequate for the comparisons studied in this paper.

2.3. The Central Conceptual Principle

Principle 2.3 (Intrinsic versus Observable Difference). Reducing resolution does not change the differences that exist between two objects. It changes the differences that can be observed or resolved.

Observed similarity may arise from hidden difference, not from absence of difference.

This principle is given precise mathematical content by the Visible–Hidden Decomposition (Section 5).

3. Axiomatization of the Layer Space

The layer space $\mathcal{L}$ requires an explicit axiomatic structure tailored to the present framework. The individual ingredients – partially ordered sets, Hilbert spaces, contractions, and intertwining maps – are classical; what is new is their specific assembly as conditions on a layer-dependent similarity structure. The six axioms below state this assembly precisely.

Definition 3.1 (Layer Space)**.** A *layer space* is a triple $(\mathcal{L}, \preceq_{\mathcal{L}}, \phi)$ where:

(L1) *Partial order:* $(\mathcal{L}, \preceq_{\mathcal{L}})$ is a partially ordered set. The relation $\ell_1 \preceq_{\mathcal{L}} \ell_2$ is read "ℓ_1 is a coarser layer than ℓ_2".

(L2) *Join existence:* For any $\ell_1, \ell_2 \in \mathcal{L}$, a least upper bound $\ell_1 \vee \ell_2$ exists in $\mathcal{L}$. Since $\ell_1 \preceq_{\mathcal{L}} \ell_2$ means ℓ_1 is *coarser* than ℓ_2 (i.e. ℓ_2 is finer), an *upper bound* u satisfies $\ell_1 \preceq_{\mathcal{L}} u$ and $\ell_2 \preceq_{\mathcal{L}} u$, meaning u is finer than or equal to both. The *least* upper bound $\ell_1 \vee \ell_2$ is therefore the *coarsest* layer that is still at least as fine as both ℓ_1 and ℓ_2. Intuitively, $\ell_1 \vee \ell_2$ is the most parsimonious descriptive layer that retains all information present in either ℓ_1 or ℓ_2.

(L3) *Description map:* There exists a map $\phi : X \times \mathcal{L} \to \mathcal{F}$, where $\mathcal{F}$ is a fixed family of feature spaces, such that $\phi(A, \ell)$ is the description of object A at layer ℓ. For $\ell_1 \preceq_{\mathcal{L}} \ell_2$, there exists a surjective projection $\pi_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ with $\phi(A, \ell_1) = \pi_{\ell_1,\ell_2}(\phi(A, \ell_2))$ (coarser layers are images of finer layers under information-reducing projections).

(L4) *Separability:* If $\phi(A, \ell) = \phi(B, \ell)$ for every $\ell \in \mathcal{L}$, then $A = B$. The collection of all layers jointly distinguishes all objects.

(L5) *Contractive layer coarsening:* Each feature space $\mathcal{F}_\ell$ is a Hilbert space (with norm $\|\cdot\|_\ell$), and for every pair $\ell_1 \preceq_{\mathcal{L}} \ell_2$ the projection $\pi_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ is a *contraction*: for all $u, v \in \mathcal{F}_{\ell_2}$,

$$\|\pi_{\ell_1,\ell_2}(u) - \pi_{\ell_1,\ell_2}(v)\|_{\ell_1} \;\leq\; \|u - v\|_{\ell_2}\,.$$

This holds automatically when π_{ℓ_1,ℓ_2} is realized as an orthogonal projection between Hilbert spaces.

(L6) *Layer–Resolution Compatibility:* Each feature space $\mathcal{F}_\ell$ from axiom (L5) is realized as a closed subspace of a common ambient Hilbert space H. For each $r \in \mathcal{R}$, let $R_r^{(\ell)} : \mathcal{F}_\ell \to \mathcal{F}_\ell$ denote the *restricted resolution operator*, i.e. the compression of a global operator R_r on H to the subspace $\mathcal{F}_\ell$. For every pair $\ell_1 \preceq_{\mathcal{L}} \ell_2$ and every $r \in \mathcal{R}$, the restricted operators satisfy the *intertwining relation*

$$R_r^{(\ell_1)} \circ \pi_{\ell_1,\ell_2} \;=\; \pi_{\ell_1,\ell_2} \circ R_r^{(\ell_2)} \quad (\text{maps } \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}). \tag{2}$$

This is well-typed: the left side first applies π ($\mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$) then $R_r^{(\ell_1)}$ (on $\mathcal{F}_{\ell_1}$); the right side first applies $R_r^{(\ell_2)}$ (on $\mathcal{F}_{\ell_2}$) then π. Both sides map $\mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$.

Remark 3.2. Axiom (L1) does not require $\mathcal{L}$ to be totally ordered. The layers atomic, molecular, cellular, anatomical, physiological, behavioral, cognitive, and social are proposed as elements of $\mathcal{L}$ in biological applications. Axiom (L2) ensures that no pair of layers is entirely incomparable. Axiom (L3) is the layer analogue of the resolution operator (Definition 4.1). Axiom (L4) prevents trivial layer spaces in which objects cannot be distinguished at all. Axiom (L5) is the metric condition required for Step 1 of Theorem 7.3: surjectivity (L3) alone does not imply that coarsening reduces norms of differences. Axiom (L6) is the *intertwining compatibility condition* required for Step 2: without it, $\|u_1\| \leq \|u_2\|$ does not imply $\|R_r u_1\| \leq \|R_r u_2\|$ for an arbitrary projection R_r, because projections are not monotone with respect to the norm ordering in general. Axioms (L5) and (L6) are independent of each other and of axioms (L1)–(L4).

Remark 3.3 (What axioms (L1)–(L6) do not specify). The measure ν over $\mathcal{L}$, required for all-layer scalar aggregation, is *not* determined by axioms (L1)–(L6). It must be supplied by the application domain. This is an intentional design choice: different applications assign different weights to different layers, and the framework should not force a canonical choice.

4. Resolution: Operators, Partitions, and Lattice

4.1. Resolution Operators

Definition 4.1 (Resolution Operator)**.** A *resolution operator* at level r is a map $R_r : X \to X_r$, where X_r is the image of X under observation at resolution r. For a partial order $\preceq$ on $\mathcal{R}$ ("$r_1 \preceq r_2$" means r_1 is coarser than r_2), the family $\{R_r\}$ satisfies the *information-loss condition*:

$$r_1 \preceq r_2 \implies R_{r_1} \circ R_{r_2} = R_{r_1}. \tag{3}$$

Once a distinction has been eliminated by coarse-graining, no further coarse-graining can restore it.

Similarity at resolution r relative to a base similarity S_0 on the object space is $S_r(A, B) = S_0(R_r(A), R_r(B))$.

4.2. Hilbert Space Realization

Definition 4.2 (Resolution-Dependent Distance)**.** Let $E_A, E_B \in H$ for a Hilbert space H. Define

$$D_r(A, B) = \|R_r(E_A - E_B)\|_H ,$$

and set $S_r(A, B) = \Phi(D_r(A, B))$ for any decreasing $\Phi : [0, \infty) \to [0, 1]$ with $\Phi(0) = 1$.

Proposition 4.3 (Monotonicity under Nested Projections)**.** *If the resolution operators are orthogonal projections in H with $\mathrm{Ran}(R_{r_1}) \subseteq \mathrm{Ran}(R_{r_2})$ whenever $r_1 \preceq r_2$, then $D_{r_1}(A, B) \leq D_{r_2}(A, B)$, and for decreasing Φ, $S_{r_1}(A, B) \geq S_{r_2}(A, B)$.*

Proof. By hypothesis, $R_{r_1} = P_1$ and $R_{r_2} = P_2$ are orthogonal projections with $\mathrm{Ran}(P_1) \subseteq \mathrm{Ran}(P_2)$. For any $v \in H$, $P_1 v = P_1 P_2 v$ (since P_1 projects onto a subspace of $\mathrm{Ran}(P_2)$), so $\|P_1 v\| \leq \|P_2 v\|$ by the contractivity of orthogonal projections. Setting $v = E_A - E_B$ gives $D_{r_1} \leq D_{r_2}$. The monotonicity of S follows from the monotone decrease of Φ. □

Remark 4.4. Proposition 4.3 does not hold for general resolution operators or for arbitrary similarity functions. The nested-projection hypothesis is a *sufficient* condition, not a universal result.

4.3. Resolution as Partition and Lattice

Definition 4.5 (Resolution as Partition)**.** An observation system $O : X \to Y$ induces the equivalence $x \sim_O y \iff O(x) = O(y)$ and the partition $\mathcal{P}_O = X/\!\sim_O$. The resolution of O is identified with $\mathcal{P}_O$. Finer partitions represent higher-resolution observation systems.

Definition 4.6 (Resolution Lattice)**.** The family of all partitions of X, ordered by refinement, forms a complete lattice $(\mathrm{Res}(X), \preceq, \wedge, \vee)$. For observation systems O_1, O_2, the combined system $O_{12}(x) = (O_1(x), O_2(x))$ yields $[x]_{12} = [x]_1 \cap [x]_2$ (meet in the lattice).

5. Visible and Hidden Difference

5.1. The Decomposition

Let $\Delta E_{AB} = E_A - E_B$ for $E_A, E_B \in H$. In a Fourier (spectral) representation, $\Delta E_{AB}(x) = \int \Delta \widehat{E}_{AB}(k) e^{ikx} dk$.

Definition 5.1 (Total, Visible, and Hidden Difference)**.**

$$D_{\text{total}} = \int \left|\Delta \widehat{E}_{AB}(k)\right|^2 dk,$$

$$D_{\text{vis}}(\Lambda) = \int_{|k| \leq \Lambda} \left|\Delta \widehat{E}_{AB}(k)\right|^2 dk,$$

$$D_{\text{hid}}(\Lambda) = \int_{|k| > \Lambda} \left|\Delta \widehat{E}_{AB}(k)\right|^2 dk.$$

Principle 5.2 (Resolution Difference Decomposition).

$$\boxed{D_{\text{total}} = D_{\text{vis}}(\Lambda) + D_{\text{hid}}(\Lambda).}$$

Changing the resolution cutoff Λ redistributes total difference between visible and hidden; it does not change the total.

Proposition 5.3. *For $\Lambda_1 \le \Lambda_2$: $D_{\text{vis}}(\Lambda_1) \le D_{\text{vis}}(\Lambda_2)$ and $D_{\text{hid}}(\Lambda_1) \ge D_{\text{hid}}(\Lambda_2)$.*

5.2. Difference Spectrum and Cumulative Functions

Definition 5.4 (Difference Spectrum)**.** $\mathcal{D}_{AB}(k) = |\Delta\widehat{E}_{AB}(k)|^2$. Normalized: $p^{\Delta}_{AB}(k) = \mathcal{D}_{AB}(k)/D_{\text{total}}$.

Definition 5.5 (Cumulative Accessible Difference and Hidden Fraction)**.**

$$F_{AB}(\Lambda) = \int_{|k|\le\Lambda} p^{\Delta}_{AB}(k)\, dk, \qquad H_{AB}(\Lambda) = 1 - F_{AB}(\Lambda).$$

F_{AB} is the fraction of total difference accessible at cutoff Λ; it is non-decreasing from 0 to 1. The Resolution–Similarity Surface (Section 10.1) is built on F_{AB}.

6. Distinction Space and Refinement Invariance

Definition 6.1 (Distinction Space)**.** A *distinction space* is a measure space (Ω, μ) together with a map $\delta_{AB} : \Omega \to [0, 1]$ for each pair (A, B), where $\delta_{AB}(\omega) = 0$ denotes identity with respect to distinction ω.

Definition 6.2 (Accessible Difference and Relative Density)**.** For accessible distinctions $\Omega_r \subseteq \Omega$ at resolution r:

$$\begin{aligned} M_r &= \mu(\Omega_r), \\ D_r(A, B) &= \int_{\Omega_r} \delta_{AB}(\omega)\, d\mu(\omega), \\ V_r(A, B) &= D_r(A, B)/M_r, \quad S_r(A, B) = 1 - V_r(A, B). \end{aligned}$$

Remark 6.3 (Non-Monotonicity of V_r). While D_r is monotone in r (adding accessible distinctions cannot decrease absolute difference), the normalized quantity V_r is *not* generally monotone. Finer resolution may reveal many similar features (decreasing V_r) or many different features (increasing V_r). The framework retains M_r, D_r, and V_r as separately meaningful quantities.

Definition 6.4 (Refinement Invariance)**.** A distinction F with $\mu(F) = w$ is subdivided into $F_1, \ldots, F_n$ with $\sum_i \mu(F_i) = w$. If $w\,\delta_F = \sum_i \mu(F_i)\delta_{F_i}$, then D_r is unchanged by the subdivision. This ensures that the similarity score does not depend on the granularity of the representation chosen by the analyst.

7. Multi-Layer Similarity and the Main Theorem

7.1. Single-Layer and Multi-Layer Similarity

Definition 7.1 (Single-Layer Similarity)**.** For fixed $\ell \in \mathcal{L}$ and $r \in \mathcal{R}$, $S(A, B; \ell, r)$ is the similarity of A and B at layer ℓ and resolution r.

Definition 7.2 (Layered Similarity Profile)**.**

$$\mathbf{S}_{A,B}(r) = \big\{S(A, B; \ell, r) : \ell \in \mathcal{L}\big\}.$$

The Layered Similarity Profile is the *primary* quantitative object of the framework. It preserves all layer-specific information and avoids premature scalar aggregation.

7.2. The Layer–Resolution Monotonicity Theorem

The following theorem is the first result in this framework that simultaneously involves the layer coordinate, the resolution coordinate, and the similarity function. It is not derivable from any single one of these components alone.

Theorem 7.3 (Layer–Resolution Monotonicity)**.** *Let $(\mathcal{L}, \preceq_{\mathcal{L}}, \phi)$ be a layer space satisfying axioms* (L1)–(L6) *(Definition 3.1), and let $\{R_r\}_{r\in\mathcal{R}}$ be a family of resolution operators satisfying the information-loss condition and realized as nested orthogonal projections on the common Hilbert space H. Suppose the similarity function takes the form $S(A, B; \ell, r) = \Phi(D_r^\ell(A, B))$ where $\Phi : [0, \infty) \to [0, 1]$ is strictly decreasing, continuous, and satisfies $\Phi(0) = 1$ (which is required by axiom* (S1)*: since $D_r^\ell(A, A) = 0$ for all A, consistency with self-similarity demands $S(A, A; \ell, r) = \Phi(0) = 1$), and*

$$D_r^\ell(A, B) = \big\|R_r^{(\ell)}\big(\phi(A, \ell) - \phi(B, \ell)\big)\big\|_\ell,$$

where $\|\cdot\|_\ell$ denotes the norm on $\mathcal{F}_\ell$ and $R_r^{(\ell)}$ is the restricted operator from axiom (L6)*.*

If $\ell_1 \preceq_{\mathcal{L}} \ell_2$ and $r_1 \preceq r_2$, then

$$\boxed{S(A, B; \ell_1, r_1) \;\geq\; S(A, B; \ell_2, r_2).} \tag{4}$$

That is, coarsening both layer and resolution simultaneously cannot decrease similarity.

Proof. We establish the inequality $D_{r_1}^{\ell_1}(A, B) \leq D_{r_2}^{\ell_2}(A, B)$ in two steps.

Step 1 (Layer coarsening). By axiom *(L3)*, $\phi(A, \ell_1) = \pi_{\ell_1,\ell_2}(\phi(A, \ell_2))$, so $\phi(A, \ell_1) - \phi(B, \ell_1) = \pi_{\ell_1,\ell_2}(\phi(A, \ell_2) - \phi(B, \ell_2))$. By axiom *(L5)*, π_{ℓ_1,ℓ_2} is a contraction, hence

$$\|\phi(A, \ell_1) - \phi(B, \ell_1)\|_{\ell_1} \;\leq\; \|\phi(A, \ell_2) - \phi(B, \ell_2)\|_{\ell_2}. \tag{5}$$

(Note: the contractivity here uses (L5) directly; surjectivity from (L3) alone is not sufficient to imply this inequality.)

Step 2 (Resolution coarsening). Write $u_i = \phi(A, \ell_i) - \phi(B, \ell_i)$ for $i = 1, 2$, both viewed as elements of the common Hilbert space H. By the intertwining relation (2) of axiom *(L6)*, with $u_1 = \pi_{\ell_1,\ell_2}(u_2) \in \mathcal{F}_{\ell_1}$:

$$R_{r_2}^{(\ell_1)} u_1 = R_{r_2}^{(\ell_1)}\, \pi_{\ell_1,\ell_2}(u_2) = \pi_{\ell_1,\ell_2}\big(R_{r_2}^{(\ell_2)} u_2\big).$$

Applying the contractivity of π_{ℓ_1,ℓ_2} (L5) to $R^{(\ell_2)}_{r_2} u_2 \in \mathcal{F}_{\ell_2}$:

$$\left\|R^{(\ell_1)}_{r_2} u_1\right\|_{\ell_1} = \left\|\pi_{\ell_1,\ell_2}\left(R^{(\ell_2)}_{r_2} u_2\right)\right\|_{\ell_1} \overset{(L5)}{\leq} \left\|R^{(\ell_2)}_{r_2} u_2\right\|_{\ell_2}. \tag{6}$$

Since $R^{(\ell_1)}_{r_1}$ and $R^{(\ell_1)}_{r_2}$ are restrictions of nested orthogonal projections to $\mathcal{F}_{\ell_1}$, Proposition 4.3 gives $\left\|R^{(\ell_1)}_{r_1} u\right\|_{\ell_1} \leq \left\|R^{(\ell_1)}_{r_2} u\right\|_{\ell_1}$ for any $u \in \mathcal{F}_{\ell_1}$. Chaining all inequalities:

$$D^{\ell_1}_{r_1} = \left\|R^{(\ell_1)}_{r_1} u_1\right\|_{\ell_1} \overset{\text{Prop. 4.3}}{\leq} \left\|R^{(\ell_1)}_{r_2} u_1\right\|_{\ell_1} \overset{(6)}{\leq} \left\|R^{(\ell_2)}_{r_2} u_2\right\|_{\ell_2} = D^{\ell_2}_{r_2}.$$

Conclusion. Since Φ is decreasing and $D^{\ell_1}_{r_1} \leq D^{\ell_2}_{r_2}$,

$$S(A, B; \ell_1, r_1) = \Phi(D^{\ell_1}_{r_1}) \geq \Phi(D^{\ell_2}_{r_2}) = S(A, B; \ell_2, r_2). \qquad \square$$

Remark 7.4 (Independence of axioms *(L5)* and *(L6)*). Surjectivity (L3) does not imply contractivity: the map $\mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto 2x$, is surjective but $|2x| > |(x, y)|$ whenever $|x| > |y|/\sqrt{3}$. Hence (L5) is independent of (L3).

Contractivity (L5) does not imply the intertwining compatibility (L6): two contractions on H need not intertwine (e.g. two distinct orthogonal projections onto non-orthogonal subspaces of $\mathbb{R}^2$). Hence (L6) is independent of (L5).

Without (L6), Step 2 of the proof fails: knowing $\|u_1\|_H \leq \|u_2\|_H$ does not imply $\|R_{r_2} u_1\|_H \leq \|R_{r_2} u_2\|_H$ for an arbitrary orthogonal projection R_{r_2}, because projections are not monotone with respect to the norm ordering in general.

Corollary 7.5 (Maximum Similarity at Minimum Resolution and Layer)**.** *If $\ell_\perp$ and $r_\perp$ are the coarsest layer and resolution in $\mathcal{L}$ and $\mathcal{R}$ respectively (bottom elements, if they exist), then $S(A, B; \ell_\perp, r_\perp) \geq S(A, B; \ell, r)$ for all ℓ, r. In particular, if $S(A, B; \ell_\perp, r_\perp) < 1$, the objects remain distinguishable even at minimal resolution and minimal descriptive detail.*

8. Scalar Aggregation: When and How

The Layered Similarity Profile $\mathbf{S}_{A,B}(r)$ is a set-valued function. In many applied settings, a single number is needed. This section states precisely when and how aggregation is legitimate.

Definition 8.1 (Multi-Layer Similarity)**.** For a subset $L' \subseteq \mathcal{L}$, an aggregation operator $\mathcal{A}$, and a measure ν on $\mathcal{L}$:

$$S_{L',r}(A, B) = \mathcal{A}_{\ell \in L'}\, S(A, B; \ell, r). \tag{7}$$

The *all-layer similarity* is

$$S_{\text{all}}(A, B; r) = \frac{\int_{\mathcal{L}} S(A, B; \ell, r)\, d\nu(\ell)}{\nu(\mathcal{L})}. \tag{8}$$

Proposition 8.2 (Admissibility of the Normalised Mean)**.** *For the specific normalised-mean formula (8), the all-layer similarity $S_{\text{all}}(A, B; r)$ is a finite real-valued average in $[0, 1]$ if and only if:*

(a) Finite positive measure: $0 < \nu(\mathcal{L}) < \infty$.

(b) Measurability: $\ell \mapsto S(A, B; \ell, r)$ *is* ν*-measurable for each fixed* A*,* B*,* r*.*

Proof. **Necessity.** Equation (8) is defined here as an ordinary real-valued normalised mean: both numerator and denominator must be finite real numbers. If $\nu(\mathcal{L}) = 0$ the denominator equals 0; undefined. If $\nu(\mathcal{L}) = \infty$ the denominator equals ∞; not a finite real number. Hence $0 < \nu(\mathcal{L}) < \infty$ is necessary for the quotient to exist as a real number. Measurability of $\ell \mapsto S(\cdot; \ell, r)$ is necessary for the numerator integral to be defined in the first place.

Sufficiency. Since $0 \leq S(A, B; \ell, r) \leq 1$ for all ℓ, the integrand is bounded and ν-measurable, hence ν-integrable, and

$$0 \;\leq\; \int_{\mathcal{L}} S(A, B; \ell, r)\, d\nu(\ell) \;\leq\; \nu(\mathcal{L}) < \infty.$$

Dividing by $\nu(\mathcal{L}) \in (0, \infty)$ gives $S_{\text{all}} \in [0, 1]$. The bound $[0, 1]$ follows from $0 \leq S \leq 1$ and monotonicity of the integral; the triangle inequality is not the correct argument here.

Scope. The necessity direction applies specifically to the normalised-mean formula (8); alternative aggregations (e.g. geometric mean, or a reference measure on an extended measure space) may relax condition (a). □

Remark 8.3 (Why σ-finiteness is insufficient). A σ-finite measure satisfies $\mathcal{L} = \bigcup_n A_n$ with $\nu(A_n) < \infty$, but may still have $\nu(\mathcal{L}) = \infty$ (e.g. Lebesgue measure on $\mathbb{R}$). The quotient $\int S\, d\nu / \nu(\mathcal{L})$ then requires dividing by ∞, which does not yield a normalized average. Condition (a) is the correct condition for a well-defined weighted mean.

Remark 8.4 (Choice of ν). The measure ν is *not* determined by axioms (L1)–(L6). Different applications require different weighting of layers. In biological comparisons, an equal-weight measure over named layers is one natural choice; in cognitive science, perceptual salience provides another. The framework requires only that the chosen ν be stated and justified explicitly. Collapsing $\mathbf{S}_{A,B}(r)$ to S_{all} without specifying ν is an incomplete specification.

9. Identity as Complete Similarity

Definition 9.1 (Layer Identity)**.** $A \equiv_{\ell,r} B \iff S(A, B; \ell, r) = 1$.

Definition 9.2 (Resolution-Dependent Equivalence)**.** $A \sim_r B \iff R_r(A) = R_r(B)$. This is a genuine equivalence relation. Under nested resolution operators: $[A]_{r_{\text{fine}}} \subseteq [A]_{r_{\text{coarse}}}$. Decreasing resolution merges equivalence classes.

Definition 9.3 (All-Layer Identity)**.** Let $\mathcal{D} \subseteq \mathcal{L} \times \mathcal{R}$ be the explicitly specified domain of comparison. Objects A and B are *all-layer identical relative to* $\mathcal{D}$ if

$$S(A, B; \ell, r) = 1 \qquad \forall\, (\ell, r) \in \mathcal{D}.$$

Remark 9.4. Without specifying $\mathcal{D}$, the phrase "absolute identity at all layers and resolutions" is not mathematically well-defined. All-layer identity is always relative to a domain of comparison.

Definition 9.5 (Resolution Merge Scale)**.** $m(A, B) = \inf\{r : A \sim_r B\}$. This quantifies the minimum coarsening required before A and B become indistinguishable.

Proposition 9.6 (Ultrametric Inequality)**.** *Under nested equivalence relations,* $m(A, C) \leq \max\{m(A, B), m(B, C)\}$. *Under suitable separation conditions,* m *is an ultrametric on* X*.*

This establishes the chain: Resolution → Equivalence → Hierarchy → Ultrametric Geometry. The ultrametric construction is classical; its role here is as a derived structure within the framework of Section 2.

10. Supporting Structures

10.1. The Resolution–Similarity Surface

Definition 10.1 (Resolution–Similarity Surface)**.**

$$\mathscr{P}_{AB}(r, t) = \frac{\mu\{\omega \in \Omega_r : \delta_{AB}(\omega) \leq t\}}{\mu(\Omega_r)}, \quad t \in [0, 1].$$

The surface simultaneously encodes observational resolution r, similarity tolerance t, and the fraction of accessible structure satisfying that tolerance. It is strictly richer than $S(A, B; r)$: the scalar $S_r(A, B)$ is recovered from $\mathscr{P}_{AB}$ as a single integral, but not vice versa. For example, $\mathscr{P}_{AB}(r, 0.05) = 0.70$ means that 70% of accessible distinguishable content differs by no more than 0.05.

10.2. Statistical Resolution

Definition 10.2 (Statistical Resolution)**.** For observation system O producing output distributions P_A^O, P_B^O: $\Delta_O(A, B) = D(P_A^O, P_B^O)$, where D is a statistical divergence (e.g. total variation or KL).

Statistical resolution extends the exact partition framework to noisy channels and does not generally form the same lattice structure. Monotonicity of Δ_O in channel quality holds when channels admit a stochastic degradation ordering.

10.3. Resource-Constrained Distinguishability

Definition 10.3 (Resource-Constrained Distinguishability)**.** For resource budget R and measurement cost $C(M)$: $\Delta_R(A, B) = \sup_{C(M) \leq R} D(P_A^M, P_B^M)$.

Proposition 10.4. $R_1 \leq R_2 \implies \Delta_{R_1}(A, B) \leq \Delta_{R_2}(A, B)$.

Definition 10.5 (Critical Distinguishability Cost and Resolution Capacity)**.**

$$\begin{aligned} R_\tau(A, B) &= \inf\{R : \Delta_R(A, B) \geq \tau\}, \\ \mathfrak{R}(R) &= \sup_{C(M) \leq R} I(X; Y_M), \\ \eta(R) &= d\mathfrak{R}/dR. \end{aligned}$$

A large $R_\tau(A, B)$ indicates intrinsic difficulty of distinction. The Resolution Capacity $\mathfrak{R}(R)$ is the information about X obtainable under budget R; it is non-decreasing in R.

11. Discussion

11.1. Summary of Contributions

The present version of the framework makes four advances over the preliminary draft:

(i) Axioms (L1)–(L6) give $\mathcal{L}$ a formal mathematical structure. Axiom (L5) – Contractive Layer Coarsening – is independent of surjectivity (L3); Axiom (L6) – Layer–Resolution Compatibility – is independent of (L5) and closes the remaining gap in Step 2 of Theorem 7.3.

(ii) Theorem 7.3 is the first result that simultaneously involves layer, resolution, and similarity, and is not a consequence of any one component alone.

(iii) Proposition 8.2 states the exact conditions under which scalar aggregation of the Layered Similarity Profile is legitimate.

(iv) The framework now carries an explicit proof for the primary monotonicity result (Proposition 4.3).

11.2. Evaluation of Constructs

The following assessment organizes the constructs by their current status.

Core conceptual contribution (potentially novel as a unit). The unified object $S(A, B; \ell, r)$, the visible/hidden decomposition $D_{\text{total}} = D_{\text{vis}}(r) + D_{\text{hid}}(r)$ interpreted via Principle 2.3, and the Layered Similarity Profile $\mathbf{S}_{A,B}(r)$ constitute the distinctive proposed contribution. None of the individual components (Hilbert space projections, partition lattices, ultrametrics) is novel; their unification around the four-level hierarchy (1) is.

Mature derived structures. The spectral decomposition, partition lattice, resolution-dependent equivalence, merge scale, and resource-constrained distinguishability rest on well-developed mathematics. Their role is to provide concrete realizations of the abstract framework, not to supply its novelty.

Open axiomatic question. Axioms (L1)–(L6) formalize $\mathcal{L}$ but do not determine the measure ν over layers. This remains an open design choice that must be made by each application.

Construct to be avoided. Naïve feature counting (I_r, C_r, D_r as raw counts) is explicitly identified as inadequate: it is not invariant under subdivision of features and must not serve as the primary similarity measure.

11.3. Open Problems and Research Directions

1. *Canonical measure over layers.* Is there a construction of ν from the structure of $(\mathcal{L}, \preceq_{\mathcal{L}}, \phi)$ alone, analogous to the Haar measure on a group? A positive answer would remove the main remaining external input to the framework.

2. *Characterization of the Layered Similarity Profile.* Under what conditions does $\mathbf{S}_{A,B}(r)$, as a set-valued function of r, determine $S(A, B; \ell, r)$ for all individual ℓ? When is the profile sufficient for all comparison purposes, and when must individual layers be examined separately?

3. *Sharp conditional upward transport.* Proposition 11.6 and Theorem 11.9 together resolve the basic upward transfer question: global upward reconstruction is impossible after genuine information loss, while conditional upward transfer is possible when the resolution-visible kernel component $\|P_{\ker \pi} R_r^{(\ell_2)} u_2\| \le \varepsilon$ is controlled. The remaining open problem is to characterize the *sharpest* conditional upward bounds: which assumptions on ε and σ are both necessary and sufficient for a given quality of upward transport, and how does the bound degrade as $\varepsilon \to \infty$ (i.e. when the kernel component is entirely uncontrolled)?

11.4. Cross-Layer Transfer: Complete Asymmetric Transport Theory

We work throughout under axioms (L1)–(L6) and the Hilbert-space similarity model of Section 7. Define the *layer contraction ratio* between $\ell_1 \preceq_{\mathcal{L}} \ell_2$ as

$$\kappa(\ell_1, \ell_2) \;=\; \sup_{u \in \mathcal{F}_{\ell_2},\, u \neq 0} \frac{\|\pi_{\ell_1,\ell_2}(u)\|_{\ell_1}}{\|u\|_{\ell_2}} \;\in\; [0, 1],$$

which measures how much π_{ℓ_1,ℓ_2} compresses norms; it satisfies $\kappa(\ell_1, \ell_2) \le 1$ by axiom (L5).

Theorem 11.1 (Downward Layer Transfer)**.** *Let $\ell_1 \preceq_{\mathcal{L}} \ell_2$, let $r \in \mathcal{R}$ be fixed, and let $S(A, B; \ell, r) = \Phi(D_r^{\ell}(A, B))$ where Φ is as in Theorem 7.3 (in particular $\Phi(0) = 1$). Then*

$$D_r^{\ell_1}(A, B) \;\le\; \kappa(\ell_1, \ell_2) \cdot D_r^{\ell_2}(A, B), \tag{9}$$

and consequently

$$S(A, B; \ell_1, r) \;\ge\; \Phi\big(\kappa(\ell_1, \ell_2) \cdot D_r^{\ell_2}(A, B)\big)\,. \tag{10}$$

Proof. By axiom (L6), $D_r^{\ell_1}(A, B) = \|R_r^{(\ell_1)} u_1\|_{\ell_1}$ where $u_1 = \pi_{\ell_1,\ell_2}(u_2)$ and $u_2 = \phi(A, \ell_2) - \phi(B, \ell_2)$. By the intertwining relation (2):

$$D_r^{\ell_1}(A, B) = \|R_r^{(\ell_1)} \pi_{\ell_1,\ell_2}(u_2)\|_{\ell_1} = \|\pi_{\ell_1,\ell_2}(R_r^{(\ell_2)} u_2)\|_{\ell_1} \le \kappa(\ell_1, \ell_2) \cdot \|R_r^{(\ell_2)} u_2\|_{\ell_2} = \kappa(\ell_1, \ell_2) \cdot D_r^{\ell_2}(A, B),$$

where the inequality is the definition of $\kappa(\ell_1, \ell_2)$. Equation (10) follows since Φ is decreasing. □

Remark 11.2 (Interpretation). Theorem 11.1 is the first result in the framework that transfers information about similarity *across layers* rather than within a single layer. Equation (10) states: *high similarity at the finer layer ℓ_2 implies at least bounded similarity at the coarser layer ℓ_1, with the bound controlled by the layer contraction ratio $\kappa(\ell_1, \ell_2)$.* When $\kappa(\ell_1, \ell_2) = 0$, the coarser layer collapses all differences and the bound becomes trivially $S(A, B; \ell_1, r) \ge \Phi(0) = 1$. When $\kappa = 1$, the operator norm of π_{ℓ_1,ℓ_2} introduces no additional contraction factor in the bound, and (10) recovers exactly Theorem 7.3. Note that $\kappa = 1$ does *not* mean that π is injective or that no information is lost: an orthogonal projection onto a proper subspace satisfies $\kappa = 1$ while having a non-trivial kernel.

Remark 11.3 (Mathematical context). The key step in the proof of Theorem 11.1 is the operator-norm inequality $\|\pi u\| \le \|\pi\| \cdot \|u\|$, which is classical. The contribution is the *interpretation and integration* of $\kappa(\ell_1, \ell_2) = \|\pi_{\ell_1,\ell_2}\|$ as a *cross-layer similarity-transfer coefficient* within the Layer × Resolution framework, and the derivation of its consequences for the similarity function $S(A, B; \ell, r)$. This is the distinction that matters for peer review.

Corollary 11.4 (Similarity-to-Similarity Transfer)**.** *In the setting of Theorem 11.1, suppose additionally that* Φ *is strictly decreasing and bijective onto* $(0, 1]$*, so that* $\Phi^{-1} : (0, 1] \to [0, \infty)$ *is well-defined. Then for any pair* A, B *with* $S(A, B; \ell_2, r) > 0$*:*

$$S(A, B; \ell_1, r) \;\ge\; \Phi\Big(\kappa(\ell_1, \ell_2) \cdot \Phi^{-1}\big(S(A, B; \ell_2, r)\big)\Big). \tag{11}$$

That is, a lower bound on similarity at the finer layer ℓ_2 *implies a lower bound on similarity at the coarser layer* ℓ_1*, with the transfer controlled entirely by* $\kappa(\ell_1, \ell_2)$*.*

Proof. Since Φ is bijective, $D_r^{\ell_2}(A, B) = \Phi^{-1}(S(A, B; \ell_2, r))$. Substituting into (10) gives (11). □

Example 11.5 (Exponential kernel: closed-form transfer law)**.** Let $\Phi(d) = e^{-d}$, so $\Phi^{-1}(s) = -\ln s$ for $s \in (0, 1]$. Then $\Phi(0) = 1$ and Φ is strictly decreasing and bijective. Corollary 11.4 gives:

$$S(A, B; \ell_1, r) \ge e^{-\kappa(\ell_1,\ell_2)\cdot(-\ln S(A,B;\ell_2,r))} = S(A, B; \ell_2, r)^{\kappa(\ell_1,\ell_2)}.$$

That is:

$$\boxed{S(A, B; \ell_1, r) \;\ge\; S(A, B; \ell_2, r)^{\,\kappa(\ell_1,\ell_2)}.} \tag{12}$$

The layer contraction ratio κ acts as an *exponent*. When $\kappa \to 0$, the right side tends to 1, consistent with $\Phi(0) = 1$: the coarser layer collapses all differences. When $\kappa = 1$, the bound reduces to $S(A, B; \ell_1, r) \ge S(A, B; \ell_2, r)$. Note that $\kappa = 1$ means the *supremal norm gain* of the layer projection equals one: $\sup_{\|u\|=1} \|\pi u\| = 1$. This does *not* imply that π is an isometry on the entire unit sphere: an orthogonal projection onto a proper subspace has operator norm 1 but maps every vector with a non-zero component orthogonal to its range to a strictly shorter vector. Equality in (12) for a *specific pair* (A, B) requires additionally that the difference vector $u_{AB} = \phi(A, \ell_2) - \phi(B, \ell_2)$ itself attains the supremum, i.e. $\|\pi_{\ell_1,\ell_2}(u_{AB})\|_{\ell_1} = \|u_{AB}\|_{\ell_2}$; this is a pair-specific condition, not guaranteed by $\kappa = 1$ alone. For example: if $S(A, B; \ell_2, r) = 0.8$ and $\kappa = 0.5$, then $S(A, B; \ell_1, r) \ge 0.8^{0.5} = \sqrt{0.8} \approx 0.894$.

Proposition 11.6 (No Global Upward Transfer under Genuine Information Loss)**.** *Let* $\pi_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ *be a layer-coarsening map with* $\ker \pi_{\ell_1,\ell_2} \ne \{0\}$ *(genuine information loss). Then there exists no finite constant* $\gamma \in [1, \infty)$ *such that*

$$\|u_2\|_{\ell_2} \le \gamma \, \|\pi_{\ell_1,\ell_2}(u_2)\|_{\ell_1} \qquad \text{for all } u_2 \in \mathcal{F}_{\ell_2}. \tag{13}$$

Proof. Let $0 \ne v \in \ker \pi_{\ell_1,\ell_2}$. Then $\pi_{\ell_1,\ell_2}(v) = 0$, so the right side of (13) equals 0 while the left side equals $\|v\|_{\ell_2} > 0$. This contradicts (13) for any finite γ. □

Remark 11.7 (Interpretation: similarity reconstruction is intrinsically asymmetric under information loss). Proposition 11.6 gives the theory a central asymmetry: *downward* transfer (from fine layer to coarse) is always possible and quantified by

Theorem 11.1; *upward* transfer (from coarse layer to fine) is globally impossible after genuine information loss. Objects identical at the coarser layer may differ arbitrarily in the kernel component $u_{\ker} \in \ker \pi_{\ell_1,\ell_2}$, and no bound on $S(A, B; \ell_1, r)$ alone can control this.

To make upward transfer possible, one must bound the hidden component. The correct quantity is the *minimum gain* of π_{ℓ_1,ℓ_2} on the orthogonal complement of its kernel.

Definition 11.8 (Minimum Gain of a Layer-Coarsening Map)**.** For $\pi_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$, define the *minimum gain*

$$\sigma(\pi_{\ell_1,\ell_2}) \;=\; \inf\left\{ \frac{\|\pi_{\ell_1,\ell_2}(u)\|_{\ell_1}}{\|u\|_{\ell_2}} \;:\; u \in (\ker \pi_{\ell_1,\ell_2})^{\perp},\; u \neq 0 \right\} \in [0, \infty).$$

If $\sigma(\pi_{\ell_1,\ell_2}) > 0$, the map is bounded below on $(\ker \pi)^{\perp}$: for every $u_{\|} \in (\ker \pi_{\ell_1,\ell_2})^{\perp}$,

$$\|u_{\|}\|_{\ell_2} \leq \frac{\|\pi_{\ell_1,\ell_2}(u_{\|})\|_{\ell_1}}{\sigma(\pi_{\ell_1,\ell_2})}.$$

Note that $\sigma \leq \kappa$; in particular, equality holds when π is an isometry on $(\ker \pi)^{\perp}$ (this is a sufficient condition, not an if-and-only-if).

Theorem 11.9 (Conditional Upward Similarity Transfer)**.** *Let $\ell_1 \preceq_{\mathcal{L}} \ell_2$, $r \in \mathcal{R}$ fixed, and $S(A, B; \ell, r) = \Phi(D_r^{\ell}(A, B))$ with Φ as in Theorem 7.3. Let $u_2 = \phi(A, \ell_2) - \phi(B, \ell_2)$ and $v_2 = R_r^{(\ell_2)} u_2$. Decompose v_2 as*

$$v_2 = v_{\|} + v_{\ker}, \qquad v_{\|} \in (\ker \pi_{\ell_1,\ell_2})^{\perp},\; v_{\ker} \in \ker \pi_{\ell_1,\ell_2}.$$

Suppose $\sigma(\pi_{\ell_1,\ell_2}) > 0$ and $\|v_{\ker}\|_{\ell_2} \leq \varepsilon$ for some $\varepsilon \geq 0$ (a resolution-visible hidden-kernel bound*: the kernel component of the resolution-projected difference is small). Then:*

$$D_r^{\ell_2}(A, B) \;\leq\; \frac{D_r^{\ell_1}(A, B)}{\sigma(\pi_{\ell_1,\ell_2})} \;+\; \varepsilon, \tag{14}$$

and consequently

$$S(A, B; \ell_2, r) \;\geq\; \Phi\left(\frac{D_r^{\ell_1}(A, B)}{\sigma(\pi_{\ell_1,\ell_2})} + \varepsilon \right). \tag{15}$$

When $\varepsilon = 0$ and $\ker \pi_{\ell_1,\ell_2} = \{0\}$ (bijective coarsening), $\sigma = \|\pi_{\ell_1,\ell_2}^{-1}\|^{-1}$ and (15) together with Theorem 11.1 yields the two-sided transport law:

$$\Phi\Big(\tfrac{1}{\sigma}\Phi^{-1}(S_{\ell_1})\Big) \leq S_{\ell_2} \leq \Phi\Big(\tfrac{1}{\kappa}\Phi^{-1}(S_{\ell_1})\Big). \tag{16}$$

Proof. Set $v_2 = R_r^{(\ell_2)} u_2 = D_r^{\ell_2}$ as a vector; by axiom (L6) and the intertwining relation (2):

$$\pi_{\ell_1,\ell_2}(v_2) = \pi_{\ell_1,\ell_2}\big(R_r^{(\ell_2)} u_2\big) = R_r^{(\ell_1)} \pi_{\ell_1,\ell_2}(u_2) = R_r^{(\ell_1)} u_1,$$

where $u_1 = \pi_{\ell_1,\ell_2}(u_2)$. Hence $\|\pi_{\ell_1,\ell_2}(v_{\|})\|_{\ell_1} = \|\pi_{\ell_1,\ell_2}(v_2)\|_{\ell_1} = D_r^{\ell_1}(A, B)$ (using $\pi(v_{\ker}) = 0$). By the triangle inequality:

$$\|v_2\|_{\ell_2} \leq \|v_{\|}\|_{\ell_2} + \|v_{\ker}\|_{\ell_2}.$$

Since $v_\parallel \in (\ker \pi)^\perp$, Definition 11.8 gives $\|v_\parallel\|_{\ell_2} \le \|\pi(v_\parallel)\|_{\ell_1}/\sigma = D_r^{\ell_1}(A,B)/\sigma$. Using $\|v_{\ker}\|_{\ell_2} \le \varepsilon$ by hypothesis:

$$D_r^{\ell_2}(A,B) = \|v_2\|_{\ell_2} \le \frac{D_r^{\ell_1}(A,B)}{\sigma} + \varepsilon.$$

This is (14); (15) follows since Φ is decreasing. □

Remark 11.10 (Structural completeness of the transfer programme). Theorems 11.1, 11.9, and Proposition 11.6 together form a complete account of cross-layer similarity transport:

(i) *Downward transfer* (Theorem 11.1): always possible; controlled by κ.

(ii) *Global upward transfer* (Proposition 11.6): impossible when $\ker \pi \neq \{0\}$.

(iii) *Conditional upward transfer* (Theorem 11.9): possible when the resolution-visible hidden kernel component $\|P_{\ker \pi} R_r^{(\ell_2)} u_2\| \le \varepsilon$; controlled by σ and ε.

(iv) *Two-sided transport* (16): holds when π is bijective ($\ker \pi = \{0\}$).

Similarity transport is therefore *intrinsically asymmetric under information loss*: downward transport is always available; upward reconstruction requires either bijectivity or a separate bound on the hidden component.

11.5. Approximate Layer–Resolution Compatibility

In applications, the intertwining relation (2) of axiom (L6) will hold only approximately. The following theorem extends the downward transport bound to this setting, replacing exact compatibility with an *approximate intertwining error* δ.

Theorem 11.11 (Robust Downward Transport under Approximate Compatibility)**.** *Let all hypotheses of Theorem 11.1 hold except that axiom* (L6) *is replaced by the approximate condition: for some $\delta \ge 0$ and every $r \in \mathcal{R}$,*

$$\big\| R_r^{(\ell_1)} \pi_{\ell_1,\ell_2}(u) - \pi_{\ell_1,\ell_2} R_r^{(\ell_2)}(u) \big\|_{\ell_1} \le \delta \, \|u\|_{\ell_2} \qquad \forall u \in \mathcal{F}_{\ell_2}. \tag{17}$$

Then, for $u_2 = \phi(A,\ell_2) - \phi(B,\ell_2)$:

$$D_r^{\ell_1}(A,B) \;\le\; \kappa(\ell_1,\ell_2) \cdot D_r^{\ell_2}(A,B) \;+\; \delta \, \|u_2\|_{\ell_2}, \tag{18}$$

and consequently

$$S(A,B;\ell_1,r) \;\ge\; \Phi\big(\kappa \cdot D_r^{\ell_2}(A,B) + \delta \, \|u_2\|_{\ell_2}\big). \tag{19}$$

Proof. From axiom (L3), $u_1 = \pi_{\ell_1,\ell_2}(u_2)$, so $D_r^{\ell_1} = \|R_r^{(\ell_1)} u_1\|_{\ell_1} = \|R_r^{(\ell_1)} \pi(u_2)\|_{\ell_1}$. Adding and subtracting $\pi R_r^{(\ell_2)} u_2$:

$$\begin{aligned} D_r^{\ell_1} &= \big\| R_r^{(\ell_1)} \pi(u_2) \big\|_{\ell_1} \\ &\le \big\| \pi(R_r^{(\ell_2)} u_2) \big\|_{\ell_1} + \big\| R_r^{(\ell_1)} \pi(u_2) - \pi(R_r^{(\ell_2)} u_2) \big\|_{\ell_1} \\ &\le \kappa \cdot \|R_r^{(\ell_2)} u_2\|_{\ell_2} + \delta \|u_2\|_{\ell_2} \\ &= \kappa \cdot D_r^{\ell_2}(A,B) + \delta \|u_2\|_{\ell_2}, \end{aligned}$$

where the second inequality uses the approximate compatibility condition (17), and the first uses $\|\pi(R_r^{(\ell_2)} u_2)\|_{\ell_1} \le \|\pi\|_{\mathrm{op}} \cdot \|R_r^{(\ell_2)} u_2\|_{\ell_2} = \kappa \cdot D_r^{\ell_2}(A,B)$ (operator-norm bound; $\kappa = \|\pi\|_{\mathrm{op}}$ by Definition 3.1). Equation (19) follows since Φ is decreasing. □

Remark 11.12 (Interpretation and recovery). When $\delta = 0$, Theorem 11.11 reduces to Theorem 11.1: exact compatibility is a special case of approximate compatibility. When $\delta > 0$, the bound (18) degrades gracefully: a larger intertwining error δ, or a larger difference vector $\|u_2\|$, widens the gap between the downward bound and the true distance. In applications, δ can be estimated from a calibration set as the operator norm of $R_r^{(\ell_1)}\hat{\pi} - \hat{\pi}R_r^{(\ell_2)}$; this does not require $\delta = 0$, only that δ be small.

Corollary 11.13 (Relative Robust Transport)**.** *In the setting of Theorem 11.11, suppose additionally that the approximate intertwining error is measured* relative to the post-resolution norm*: there exists $\delta_r \geq 0$ such that for every $u \in \mathcal{F}_{\ell_2}$,*

$$\left\| R_r^{(\ell_1)}\pi(u) - \pi R_r^{(\ell_2)}(u) \right\|_{\ell_1} \leq \delta_r \, \|R_r^{(\ell_2)}u\|_{\ell_2}. \tag{20}$$

Then:

$$D_r^{\ell_1}(A, B) \;\leq\; (\kappa + \delta_r) \cdot D_r^{\ell_2}(A, B), \tag{21}$$

and, if Φ is invertible on $(0, 1]$,

$$S(A, B; \ell_1, r) \;\geq\; \Phi\big((\kappa + \delta_r)\, \Phi^{-1}(S(A, B; \ell_2, r))\big). \tag{22}$$

For the exponential kernel $\Phi(d) = e^{-d}$:

$$\boxed{S(A, B; \ell_1, r) \;\geq\; S(A, B; \ell_2, r)^{\,\kappa+\delta_r}.} \tag{23}$$

Proof. Apply the triangle inequality as in Theorem 11.11, then use (20) in place of (17):

$$D_r^{\ell_1} \leq \kappa D_r^{\ell_2} + \delta_r D_r^{\ell_2} = (\kappa + \delta_r) D_r^{\ell_2}.$$

This is (21). Equation (22) follows from the same substitution as Corollary 11.4; (23) is the exponential-kernel special case. □

Remark 11.14 (Why the relative form is preferable in practice). Theorem 11.11 requires knowing $\|u_2\|$, the norm of the full (unresolved) difference vector. In most applications, only the resolved quantity $D_r^{\ell_2} = \|R_r^{(\ell_2)}u_2\|$ is directly observable. Corollary 11.13 replaces $\delta\|u_2\|$ by $\delta_r D_r^{\ell_2}$, so the bound is expressed entirely in terms of $D_r^{\ell_2}$ (equivalently, S_{ℓ_2}). The resulting similarity law (23) $S_{\ell_1} \geq S_{\ell_2}^{\kappa+\delta_r}$ is the natural robust analogue of the exact law $S_{\ell_1} \geq S_{\ell_2}^{\kappa}$ from Corollary 11.4: the exponent κ is inflated by the relative intertwining error δ_r, and the bound degrades continuously to the exact case as $\delta_r \to 0$.

Remark 11.15 (Kernel-compatibility requirement). The relative condition (20) has an implicit structural consequence: when $R_r^{(\ell_2)}u = 0$, the right side vanishes, so the left side must also be zero, i.e. $R_r^{(\ell_1)}\pi(u) = 0$. In other words:

$$\ker R_r^{(\ell_2)} \;\subseteq\; \ker\big(R_r^{(\ell_1)}\pi_{\ell_1,\ell_2}\big).$$

This *kernel-compatibility condition* requires that information completely removed by the fine-layer resolution operator cannot reappear after layer coarsening. It is a natural information-monotonicity property and is consistent with the framework's central principle (Principle 2.3): coarsening cannot create visible information from hidden information. Applications should verify this condition on the learned maps.

Definition 11.16 (Minimal Relative Compatibility Constant)**.** For admissible maps $(\pi_{\ell_1,\ell_2}, R_r^{(\ell_1)}, R_r^{(\ell_2)})$ satisfying the kernel-compatibility condition of Remark 11.15, define

$$\delta_r^*(\ell_1,\ell_2) \;=\; \sup_{\substack{u\in\mathcal{F}_{\ell_2}\\ R_r^{(\ell_2)}u\neq 0}} \frac{\|R_r^{(\ell_1)}\pi_{\ell_1,\ell_2}(u) - \pi_{\ell_1,\ell_2}R_r^{(\ell_2)}(u)\|_{\ell_1}}{\|R_r^{(\ell_2)}u\|_{\ell_2}}.$$

The quantity δ_r^* is the smallest constant δ_r for which the relative approximate compatibility condition (20) holds. By construction, Corollary 11.13 applies with $\delta_r = \delta_r^*$, giving the tightest possible bound from that corollary:

$$S(A,B;\ell_1,r) \;\geq\; S(A,B;\ell_2,r)^{\kappa+\delta_r^*}.$$

When π and R_r are finite-dimensional linear maps, δ_r^* is the operator norm of $R_r^{(\ell_1)}\pi - \pi R_r^{(\ell_2)}$ restricted to the complement of $\ker R_r^{(\ell_2)}$, normalised by $\|R_r^{(\ell_2)}\cdot\|$; it can be computed exactly and constitutes an intrinsic compatibility quantity of the operator triple $(\pi_{\ell_1,\ell_2},\, R_r^{(\ell_1)},\, R_r^{(\ell_2)})$ relative to the chosen Hilbert-space norms. Note: in infinite dimensions, $\delta_r^* < \infty$ is not automatic; a sufficient condition is that $R_r^{(\ell_2)}$ be bounded below on $(\ker R_r^{(\ell_2)})^\perp$, or equivalently that $R_r^{(\ell_2)}$ have closed range. In finite dimensions this is always satisfied.

Proposition 11.17 (Pseudoinverse Characterization of δ_r^*)**.** *Let $\mathcal{F}_{\ell_2}$, $\mathcal{F}_{\ell_1}$ be Hilbert spaces and let $\pi = \pi_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ and $R = R_r^{(\ell_2)} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_2}$, $P = R_r^{(\ell_1)} : \mathcal{F}_{\ell_1} \to \mathcal{F}_{\ell_1}$ be bounded linear operators with R having closed range. Let $R^\dagger$ denote the Moore–Penrose pseudoinverse of R, defined on $\mathrm{Ran}(R)\oplus\mathrm{Ran}(R)^\perp$ with $R^\dagger R = P_{(\ker R)^\perp}$ (the orthogonal projection onto $(\ker R)^\perp$).*

Then

$$\delta_r^* = \left\| (P\pi - \pi R)\, R^\dagger \right\|_{\mathrm{op}}, \tag{24}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm from $\mathcal{F}_{\ell_2}$ to $\mathcal{F}_{\ell_1}$.

Proof. We first simplify the supremum in Definition 11.16. For $u \in \mathcal{F}_{\ell_2}$ with $Ru \neq 0$, write $v = Ru$; since R has closed range, every $v \in \mathrm{Ran}(R)$ is of this form with $u = R^\dagger v + w$ for some $w \in \ker R$. Since the condition $Ru \neq 0$ means $v = Ru \neq 0$:

$$P\pi(u) - \pi(Ru) = P\pi(R^\dagger v + w) - \pi v = P\pi R^\dagger v + P\pi w - \pi v.$$

By kernel-compatibility (Remark 11.15), $\ker R \subseteq \ker(P\pi)$, i.e. $P\pi w = 0$ for every $w \in \ker R$ (since $P\pi w = 0$ is exactly the condition $w \in \ker(P\pi)$). Therefore the kernel component drops: $P\pi w = 0$, and the expression reduces to $(P\pi R^\dagger - \pi)v$. We now rewrite $(P\pi R^\dagger - \pi)v$ in terms of the *intertwining-defect operator* $E := P\pi - \pi R$. Since $RR^\dagger$ is the orthogonal projection onto $\mathrm{Ran}(R)$ and $v \in \mathrm{Ran}(R)$, we have $RR^\dagger v = v$. Therefore $\pi(RR^\dagger v) = \pi v$, i.e. $(\pi R)R^\dagger v = \pi v$. Substituting:

$$P\pi R^\dagger v - \pi v = P\pi R^\dagger v - (\pi R)R^\dagger v = (P\pi - \pi R)R^\dagger v.$$

Therefore:

$$\begin{aligned}
\delta_r^* &= \sup_{\substack{u\in\mathcal{F}_{\ell_2}\\ Ru\neq 0}} \frac{\|P\pi u - \pi Ru\|_{\ell_1}}{\|Ru\|_{\ell_2}} \\
&= \sup_{0\neq v\in\mathrm{Ran}(R)} \frac{\|(P\pi - \pi R)R^\dagger v\|_{\ell_1}}{\|v\|_{\ell_2}} \\
&= \|(P\pi - \pi R)R^\dagger\|_{\mathrm{op}},
\end{aligned}$$

which is (24). □

Remark 11.18 (Computational implications). Equation (24) gives a direct route to computing δ_r^* in finite dimensions: form the matrix $E := P\pi - \pi R$ (the *intertwining-defect operator* of the triple), multiply on the right by the pseudoinverse $R^\dagger$, and take the largest singular value. In the finite-dimensional setting, this is a single SVD computation. The pseudoinverse formulation also clarifies why the kernel-compatibility condition (Remark 11.15) is needed: without it, the step $P\pi w = 0$ fails and δ_r^* would not reduce to the compact form (24).

Remark 11.19 (Relationship to exact axiom (L6)). Axiom (L6) asserts that $P\pi = \pi R$ exactly (equation (2) in operator notation). Under exact (L6), the intertwining-defect $E = P\pi - \pi R = 0$, so Proposition 11.17 immediately gives $\delta_r^* = 0$.

The converse holds under an additional assumption. $\delta_r^* = 0$ means $(P\pi - \pi R)R^\dagger = 0$, i.e. $P\pi u = \pi R u$ for all $u \in (\ker R)^\perp$. To extend this to all of $\mathcal{F}_{\ell_2}$, one needs $P\pi w = \pi R w$ for $w \in \ker R$. Since $Rw = 0$, this requires $P\pi w = 0$ for all $w \in \ker R$, i.e. $\pi(\ker R) \subseteq \ker P$. The kernel-compatibility condition (Remark 11.15) gives $\ker R \subseteq \ker(P\pi)$, i.e. $P\pi w = 0$ for $w \in \ker R$ – which is exactly what is needed. Hence:

$$\delta_r^* = 0 \iff P\pi u = \pi R u \text{ for all } u \in (\ker R)^\perp, \tag{25}$$

and under the kernel-compatibility condition, (25) is equivalent to full axiom (L6): $P\pi = \pi R$ everywhere.

Caveat. Without kernel-compatibility, the implication $\delta_r^* = 0 \Rightarrow$ (L6) may fail: it is possible that $\delta_r^* = 0$ while $P\pi w \neq 0$ for some $w \in \ker R$. Kernel-compatibility is therefore a necessary part of the equivalence, not just a technical convenience.

This provides a measurable characterisation: compute δ_r^* via SVD, verify kernel-compatibility on a calibration set, and check whether $\delta_r^* = 0$.

Lemma 11.20 (Bound on δ_r^* when R is an orthogonal projector)**.** *Suppose $R = R_r^{(\ell_2)}$ is an orthogonal projection on $\mathcal{F}_{\ell_2}$ ($R^2 = R$, $R^\dagger = R$, $\|R\| = 1$). Then*

$$\delta_r^* \leq (\|P\| + 1)\,\kappa. \tag{26}$$

If additionally $P = R_r^{(\ell_1)}$ is an orthogonal projector ($\|P\| = 1$), then $\delta_r^* \leq 2\kappa$; Corollary 11.22 (proved from Theorem 11.21) sharpens this to $\delta_r^* \leq \kappa$ without any range-inclusion assumption. If additionally $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)$, then $\delta_r^* = 0$ (Corollary 11.23).

Proof. Since $R^\dagger = R$ and $\|R\| = 1$, Proposition 11.17 gives $\delta_r^* = \|(P\pi - \pi R)R\|$. By the triangle inequality and sub-multiplicativity:

$$\|(P\pi - \pi R)R\| \leq \|P\pi R\| + \|\pi R^2\| = (\|P\| \cdot \kappa + \kappa) \cdot \|R\| = (\|P\| + 1)\kappa,$$

which is (26).

When P, R both orthogonal projectors, $\|P\| = 1$ and $\sigma_R = 1$, so (26) gives $\delta_r^* \leq 2\kappa$. The sharper bound $\delta_r^* \leq \kappa$ then follows from Corollary 11.22, which uses Theorem 11.21.

Case: $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)$ **(range inclusion).** When $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)$, P acts as the identity on $\mathrm{Ran}(\pi R)$: $P(\pi R v) = \pi R v$ for all $v \in \mathrm{Ran}(R)$. Hence for $v \in \mathrm{Ran}(R)$:

$$(P\pi - \pi R)v = P(\pi v) - \pi(Rv) \overset{Rv=v}{=} P(\pi v) - \pi v \overset{\pi v \in \mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)}{=} \pi v - \pi v = 0.$$

So $\delta_r^* = 0$ in this case (not merely $\leq \kappa$).

General case without range inclusion. The coarse bound $\delta_r^* \leq 2\kappa$ from (26) (with $\|P\| = 1$ and $\|R\| = 1$) applies. The exact value is $\delta_r^* = \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$ by Theorem 11.21. See Corollary 11.22 for the theorem-level bound $\delta_r^* \leq \kappa$ when P is also an orthogonal projector. □

Theorem 11.21 (Tight Characterisation of δ_r^*)**.** *Let both $P = R_r^{(\ell_1)}$ and $R = R_r^{(\ell_2)}$ be orthogonal projectors and suppose the kernel-compatibility condition (Remark 11.15) holds. Then*

$$\delta_r^* = \left\|(P - I)\pi\big|_{\mathrm{Ran}(R)}\right\|_{\mathrm{op}} := \sup_{v \in \mathrm{Ran}(R),\, v \neq 0} \frac{\|(P - I)\pi v\|}{\|v\|}. \tag{27}$$

Proof. **Step 1: Key identity.** For any $v \in \mathrm{Ran}(R)$, since R is an orthogonal projector $R^2 = R$, so $v = Rv$ and $\pi Rv = \pi v$. Therefore:

$$(P\pi - \pi R)v = P\pi v - \pi Rv = P\pi v - \pi v = (P - I)\pi v. \tag{28}$$

Step 2: Upper bound $\delta_r^* \leq \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$. From Proposition 11.17, $\delta_r^* = \|(P\pi - \pi R)R\|_{\mathrm{op}}$ (using $R^\dagger = R$). For any u with $Ru \neq 0$, let $v := Ru \in \mathrm{Ran}(R)$. Since $v \in \mathrm{Ran}(R)$, identity (28) gives $(P\pi - \pi R)v = (P - I)\pi v$, and $(P\pi - \pi R)Ru = (P\pi - \pi R)v = (P - I)\pi v$. Both the numerator $\|(P\pi - \pi R)Ru\| = \|(P - I)\pi v\|$ and the denominator $\|Ru\| = \|v\|$ depend on u only through $v = Ru$. Therefore the ratio $(\|(P\pi - \pi R)Ru\|)/\|Ru\|$ is a function of $v \in \mathrm{Ran}(R)$ alone, and:

$$\delta_r^* = \sup_{u:\, Ru \neq 0} \frac{\|(P\pi - \pi R)Ru\|}{\|Ru\|} = \sup_{v \in \mathrm{Ran}(R),\, v \neq 0} \frac{\|(P - I)\pi v\|}{\|v\|},$$

giving $\delta_r^* \leq \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$.

Step 3: Lower bound $\delta_r^* \geq \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$. For any $v \in \mathrm{Ran}(R)$, $v \neq 0$, take $u := v$. Since $v \in \mathrm{Ran}(R)$, $Rv = v \neq 0$, so $u = v$ is admissible in the supremum of Definition 11.16 (the admissibility condition $Ru \neq 0$ holds: $Ru = v \neq 0$, since $v \in \mathrm{Ran}(R)$ and $v \neq 0$). The numerator and denominator of Definition 11.16 evaluated at $u = v$:

$$\begin{aligned} \text{numerator:} \quad & \|P\pi(Ru) - \pi R(Ru)\| = \|P\pi v - \pi Rv\| \overset{(28)}{=} \|(P - I)\pi v\|, \\ \text{denominator:} \quad & \|Ru\| = \|Rv\| = \|v\|. \end{aligned}$$

Hence $\delta_r^* \geq \|(P - I)\pi v\|/\|v\|$. Taking the sup over all $v \in \mathrm{Ran}(R)$, $v \neq 0$, gives $\delta_r^* \geq \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$.

Combining Steps 2 and 3 gives (27). □

Corollary 11.22 (Sharp projector bound)**.** *Under the hypotheses of Theorem 11.21 (P, R orthogonal projectors, kernel-compatibility),*

$$0 \leq \delta_r^* \leq \kappa. \tag{29}$$

Proof. By Theorem 11.21, $\delta_r^* = \|(P - I)\pi|_{\mathrm{Ran}(R)}\|$. Since P is an orthogonal projector, $I - P$ is also an orthogonal projector, so $\|P - I\| = \|I - P\| \leq 1$. Therefore:

$$\delta_r^* = \|(P - I)\pi|_{\mathrm{Ran}(R)}\| \leq \|P - I\| \cdot \|\pi|_{\mathrm{Ran}(R)}\| \leq 1 \cdot \|\pi\| = \kappa.$$

□

Corollary 11.23 (Zero-defect characterisation)**.** *Under the hypotheses of Theorem 11.21,*

$$\delta_r^* = 0 \quad \Longleftrightarrow \quad \mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P). \tag{30}$$

Proof. By Theorem 11.21, $\delta_r^* = \|(P-I)\pi|_{\mathrm{Ran}(R)}\| = 0$ if and only if $(P-I)\pi v = 0$ for all $v \in \mathrm{Ran}(R)$. Since P is an orthogonal projector, $(P-I)y = 0 \iff y \in \mathrm{Ran}(P)$. Hence $(P-I)\pi v = 0$ for all $v \in \mathrm{Ran}(R)$ if and only if $\pi v \in \mathrm{Ran}(P)$ for all $v \in \mathrm{Ran}(R)$, i.e. $\pi(\mathrm{Ran}(R)) \subseteq \mathrm{Ran}(P)$, which is $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)$. □

Remark 11.24 (Role of kernel-compatibility). The theorem's three steps use kernel-compatibility differently.

Step 1 uses only $R^2 = R$: $\pi R v = \pi(Rv) = \pi v$ for $v \in \mathrm{Ran}(R)$. No kernel-compatibility required.

Step 2 invokes Proposition 11.17, which was proved under kernel-compatibility ($\ker R \subseteq \ker(P\pi)$, Remark 11.15). That condition ensures the kernel component $w \in \ker R$ drops from $(P\pi - \pi R)Ru$ in the derivation of Prop. 11.17, making the pseudoinverse formula $\delta_r^* = \|(P\pi - \pi R)R^\dagger\|$ valid. Without it, Prop. 11.17 may not hold, and the starting equality $\delta_r^* = \|(P\pi - \pi R)R\|$ in Step 2 would be unwarranted.

Step 3 takes $u = v \in \mathrm{Ran}(R)$ directly. Admissibility ($Ru = v \neq 0$) holds without kernel-compatibility. The equality $(P\pi - \pi R)v = (P - I)\pi v$ follows from Step 1 (i.e. from $R^2 = R$). No further condition is needed.

Remark 11.25 (Degenerate training regime). When $\delta_r^* \approx \kappa$, the robust bound gives $S_{\ell_1} \geq S_{\ell_2}^{\kappa+\delta_r^*} \approx S_{\ell_2}^{2\kappa}$. Since $2\kappa > \kappa$, the lower bound on S_{ℓ_1} is smaller – the bound degrades, which is the correct qualitative behaviour. Lemma 11.20 shows the inflation is at most a factor of 2 when both P, R are orthogonal projectors. Practical training should aim for $\delta_r^* \ll \kappa$.

Example 11.26 (Computed δ_r^* with orthogonal projectors)**.** Let $\mathcal{F}_{\ell_2} = \mathbb{R}^3$, $\mathcal{F}_{\ell_1} = \mathbb{R}^2$, and set:

$$\pi = \begin{pmatrix} 3/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix} \quad \text{(the map from Example B.5)},$$

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{(orthogonal proj. onto first two coords)},$$

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_{2\times 2} \quad \text{(identity on } \mathcal{F}_{\ell_1}\text{)}.$$

Since $P = I$, the intertwining-defect operator $E = P\pi - \pi R = \pi - \pi R = \pi(I - R)$.

Computing δ_r^*. By Proposition 11.17 (with $R^\dagger = R$):

$$\delta_r^* = \|(\pi - \pi R)R\|_{\mathrm{op}} = \|\pi(I - R)R\|_{\mathrm{op}} = \|\pi \cdot 0\| = 0,$$

since $(I - R)R = 0$ (R is a projector). This confirms: $P = I$ means $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(I) = \mathcal{F}_{\ell_1}$ trivially, so the range-inclusion condition of Lemma 11.20 holds, giving $\delta_r^* = 0$.

Adding a perturbation. Suppose now $P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ (projection onto first coordinate), π and R as above. Then:

$$P\pi = \begin{pmatrix} 3/4 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \pi R = \begin{pmatrix} 3/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix},$$
$$P\pi - \pi R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1/2 & 0 \end{pmatrix}.$$

With $R^\dagger = R$:

$$(P\pi - \pi R)R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1/2 & 0 \end{pmatrix}.$$

The operator norm (largest singular value) is $\delta_r^* = 1/2$. Since $\kappa = \|\pi\| = 3/4$, we have $\delta_r^* = 1/2 < 3/4 = \kappa$, consistent with Lemma 11.20 ($\delta_r^* = 1/2 < 3/4 = \kappa$, so the 2κ bound gives $\delta_r^* \le 3/2$, and indeed $1/2 < 3/2$). The range-inclusion check: $\mathrm{Ran}(\pi R) \subseteq \mathrm{Ran}(P)$ requires that all images of πR lie in the first coordinate axis. $\pi R = \begin{pmatrix} 3/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix}$, whose column span includes $(0, 1/2, 0)^T$, outside $\mathrm{Ran}(P) = \{(x, 0)^T\}$. So the range inclusion fails. Lemma 11.20 (without range inclusion) gives $\delta_r^* \le 2\kappa = 3/2$ as a certified upper bound. The exact value is given by Theorem 11.21: $\delta_r^* = \|(P - I)\pi|_{\mathrm{Ran}(R)}\| = \| \begin{pmatrix} 0 & 0 \\ 0 & -1/2 \end{pmatrix} \| = 1/2$, which lies strictly below $2\kappa = 3/2$. The bound $\delta_r^* \le \kappa = 3/4$ is a theorem (Corollary 11.22), since both P and R are orthogonal projectors; indeed $1/2 \le 3/4$. The strict inequality $\delta_r^* = 1/2 < 3/4 = \kappa$ is specific to the chosen operators and does not hold in general without range inclusion. The robust similarity bound: $S_{\ell_1} \ge S_{\ell_2}^{3/4+1/2} = S_{\ell_2}^{5/4}$.

Remark 11.27 (Applicability to learned maps). Theorem 11.11 is the key theoretical result enabling the LLM and computer-vision applications (Sections 13, C). Learned admissible maps (approach (A2)) will satisfy (L6) only approximately; Theorem 11.11 and Corollary 11.13 guarantee that the transport bound remains valid up to the intertwining error δ or δ_r. Strengthening the compatibility regularizer in the training loss *encourages* δ_r to decrease, subject to the representation-fidelity trade-off: a very strong regularizer may force $\delta_r \to 0$ at the cost of degraded representation quality. In practice, δ_r should be measured on a held-out set rather than assumed to be negligible.

11.6. Relation to Existing Literature

The asymmetric transport structure {downward, upward impossibility, conditional upward} developed in Sections 11.4–11.9 is related to, but distinct from, several bodies of existing work. This subsection positions the framework relative to the closest antecedents.

Data Processing Inequality and Blackwell Ordering. The Data Processing Inequality (DPI) in information theory [Cover and Thomas, 2006] states that applying a stochastic channel to a pair of distributions cannot increase their Kullback–Leibler divergence or mutual information. This is a *downward* result in the same spirit as Theorem 11.1: coarser observation cannot increase the distinguishability of two objects. Blackwell's theorem on comparison of statistical experiments gives a partial order on channels (experiments) by their distinguishability power, and a channel O_1 is Blackwell-dominated by O_2 if and only if every decision problem is at least as well-solved by O_2. The present framework can be viewed as a *deterministic* and *multi-layer* analogue: the layer coarsening map π_{ℓ_1,ℓ_2} plays the role of a deterministic channel, κ quantifies its information-reducing effect on norms, and Proposition 11.6 is the analogue of the irreversibility of information loss in the DPI. The distinguishing feature of the present work is the simultaneous treatment of *layer* and *resolution* as independent parameters governing similarity.

Operator-theoretic antecedents. The contraction ratio $\kappa = \|\pi\|$ and minimum gain σ are instances of the operator norm and the minimum modulus (or lower bound of the spectrum) of a bounded linear operator [Reed and Simon, 1980, Halmos, 1982]. The factorization $u_2 = u_{\|} + u_{\ker}$ used in Theorem 11.9 is the standard orthogonal decomposition relative to the kernel; the key novelty is the interpretation of $\|u_{\ker}\|$ as a quantitative measure of *hidden layer-difference* that must be bounded for upward reconstruction.

Coarse-graining and renormalization. In statistical physics and quantum field theory, coarse-graining (renormalization group flow) is likewise asymmetric: a block-spin transformation reduces the number of degrees of freedom and the transformation is in general non-invertible. Ultrametric clustering (Section 9) is a deterministic analogue [Rammal et al., 1986]. The present framework adds to this picture an explicit *similarity function* that tracks what is preserved and what is lost under coarsening, via the visible/hidden decomposition.

Representation transfer and transfer learning. In machine learning, *transfer learning* moves representations trained on a source task to a target task. The layer-to-layer transfer studied here is structurally different: it concerns the relationship between similarity values computed at different levels of abstraction over the *same* objects, rather than the reuse of learned parameters across tasks. The metric-learning literature [Kulis, 2013] studies conditions under which distance functions generalise across tasks, but does not incorporate the layer/resolution double parameter or the visibility decomposition.

Scale-space and multiscale similarity. Scale-space theory [Witkin, 1983, Lindeberg, 1994] studies how image representations change as Gaussian blur is applied at increasing scale. The scale parameter in scale-space corresponds to the resolution parameter r in the present framework, and the monotone loss of fine structure as scale increases mirrors the visible/hidden difference decomposition. The key distinction is that scale-space operates on a single fixed representational layer (image intensity), whereas the present framework treats *layer* and *resolution* as independent axes,

and introduces a formal notion of *similarity* between objects at each point of the Layer × Resolution grid.

Blackwell's comparison of statistical experiments. Blackwell [Blackwell, 1953] showed that one statistical experiment is at least as informative as another if and only if the former can be converted to the latter by a Markov kernel. This induces a partial order on channels that is the direct analogue of the resolution lattice (Section 4), and the irreversibility of information loss under Blackwell ordering corresponds to Proposition 11.6. The present framework differs by (a) pairing the channel order with an explicit layer structure, (b) defining a quantitative similarity function $S(A, B; \ell, r)$ rather than a binary informativeness relation, and (c) deriving metric bounds $(\kappa, \sigma, \varepsilon)$ on the similarity transport.

Tversky's feature-based similarity. Tversky's seminal model [Tversky, 1977] defines similarity as a weighted combination of shared and unshared features. This is closely related to the distinction space (Ω, μ) of Section 6: the local difference function $\delta_{AB}(\omega)$ generalises Tversky's feature indicators, and refinement invariance (Section 6) addresses the feature-weighting arbitrariness that Tversky's model leaves open. The present framework adds the layer and resolution dimensions that Tversky's model does not contain.

Bounded-below operators and minimum modulus. The minimum gain $\sigma(\pi)$ defined in Definition 11.8 is equivalent to the minimum modulus of the operator π on $(\ker \pi)^\perp$, a classical object in functional analysis [Kato, 1995]. For a bounded linear operator between Hilbert spaces, being bounded below on $(\ker \pi)^\perp$ (i.e. $\sigma > 0$) is equivalent to its range being closed [Kato, 1995]; when $\mathcal{F}_{\ell_2}$ is finite-dimensional, both conditions hold automatically. The novelty in the present context is not the definition of σ itself but its role as the precise quantity governing how much similarity information can be recovered from a coarser layer when the hidden kernel component is bounded.

Summary of positioning. To the best of the author's knowledge, the specific combination of

(i) a two-parameter (ℓ, r) similarity function;

(ii) a visible/hidden decomposition of difference;

(iii) quantitative downward bounds via κ;

(iv) a provable impossibility of global upward reconstruction under genuine coarsening;

(v) conditional upward bounds via σ and a resolution-visible kernel bound;

(vi) robust similarity-transport bounds (Theorem 11.11, Corollary 11.13) under approximate layer–resolution compatibility ($\|R_r \pi u - \pi R_r u\| \le \delta_r \|R_r u\|$), giving $S_{\ell_1} \ge S_{\ell_2}^{\kappa+\delta_r}$ and bridging the exact theory to learned representations;

does not appear in information theory, operator theory, scale-space theory, or representation learning in the unified form developed here. Each ingredient is classical; the proposed contribution is their assembly into a coherent theory of *similarity transport across descriptive layers, exact and approximate.*

11.7. Structured Comparison with Closest Antecedents

The following table organises the relationship between this framework and the bodies of work most directly related to it. The goal is not to claim that any antecedent is incorrect or superseded, but to locate precisely what is added and what is shared.

Related work	What it shares with this framework	What this framework adds
Tversky (1977) [Tversky, 1977]	Similarity as a function of shared and unshared features; the idea that description choice affects the similarity value.	Layer ℓ and resolution r as two independent, formally axiomatised parameters; directional transport bounds between layers; the visible/hidden decomposition; invariance under feature-space refinement.
Scale-space [Witkin, 1983, Lindeberg, 1994]	Representation changes continuously with observational scale; finer details disappear at coarser scale.	A second independent axis (*descriptive layer* ℓ, not merely blur scale); pair-specific similarity transport bounds; the asymmetric transport structure.
Data Processing Inequality [Cover and Thomas, 2006]	Coarsening cannot increase mutual information or KL-divergence; information loss is irreversible.	A two-parameter similarity grid instead of a scalar channel; pair-specific bounds $S_{\ell_1} \geq S_{\ell_2}^{\kappa}$ and $S_{\ell_1} \geq S_{\ell_2}^{\kappa+\delta_r^*}$; the Layered Similarity Profile.
Blackwell ordering [Blackwell, 1953]	A partial order on observation systems by information content; information loss induces a directed structure on experiments.	Quantitative similarity transport coefficients $(\kappa, \sigma, \varepsilon, \delta_r^*)$ rather than a binary/partial-order informativeness relation; conditional upward recovery.
Repr. similarity / CKA [Kornblith et al., 2019, Klabunde et al., 2025]	Comparison of neural representations across layers or models; measures of alignment between activation spaces.	Directional transport bounds; asymmetric upward/downward structure; resolution as an independent axis; the kernel obstruction; approximate compatibility via δ_r^*.
Repr. alignment / stitching [Insulla et al., 2025]	Learning maps between representation spaces; theoretical conditions for alignment.	The resolution parameter r as an independent axis; asymmetric transport; the visible/hidden difference; approximate compatibility bounds.

Related work	What it shares with this framework	What this framework adds
Multi-Level OT [Shah and Khosla, 2026]	Transport-theoretic framework operating on neural layers; layer-to-layer couplings.	Different notion of "transport": Multi-Level OT moves representation *mass*; this framework transports *pairwise similarity inequalities* under information loss.
Operator theory [Reed and Simon, 1980, Halmos, 1982, Kato, 1995]	Operator norm $\kappa = \|\pi\|$, minimum modulus σ, kernel decomposition, perturbation bounds.	All of these are classical; this framework *interprets* them as similarity-transport coefficients and combines them with the Layer× Resolution grid and the pair-specific asymmetric laws.
Platonic Projection Structures (PPS, 2026) [Ishii et al., 2026]	Operator-induced observability of latent representations; kernel as hidden information; quotient geometry; approximate intertwining ($\Phi\Pi_T \approx \Pi_S\Phi$).	PPS centres on *observability* of a latent space by a single observation operator. This framework centres on *pairwise similarity transport* across a two-dimensional Layer× Resolution grid, including the Layered Similarity Profile, the visible/hidden difference, the asymmetric downward/upward structure, and the intrinsic compatibility quantity δ_r^* of the operator triple. PPS is the closest antecedent found; the two frameworks are complementary rather than redundant.

Remark 11.28 (What the comparison shows). No single antecedent simultaneously contains all of: (a) a Layer×Resolution similarity function $S(A, B; \ell, r)$; (b) exact downward bounds $S_{\ell_1} \geq S^{\kappa}_{\ell_2}$; (c) a provable upward-reconstruction obstruction (Proposition 11.6); (d) conditional upward recovery with σ and ε; and (e) the robust relative law $S_{\ell_1} \geq S^{\kappa+\delta_r^*}_{\ell_2}$ with the intrinsically defined δ_r^*. PPS [Ishii et al., 2026] is the most structurally similar work; it and this framework should be read together, as they address complementary questions about operator-structured representation spaces.

12. Proposed Research Direction: Analogy Discovery in Mathematics

This section proposes a second line of application for the framework, distinct from computer vision. It is presented as a *research direction* rather than a worked-out result: no theorem is claimed; the purpose is to identify how the structures

already developed map onto the problem of discovering and qualifying mathematical analogies.

12.1. Mathematical Objects as Elements of the Framework

Let X be a collection of mathematical problems, theorems, or structures. A problem $P \in X$ can be described at multiple *mathematical layers*, for example:

$\mathcal{L}_{\text{math}} = \{\text{algebraic, geometric, topological, combinatorial, spectral, probabilistic}\}$.

The description map $\phi(P, \ell)$ assigns to each problem its representation in layer ℓ (e.g. its algebraic structure, its spectral invariants, its combinatorial encoding). Resolution r within a layer corresponds to the degree of detail retained: exact description, finite quotient, asymptotic expansion, low-order invariants, and so on.

12.2. Similarity as Analogy Detection

Given an unsolved problem P and a solved problem Q, the function $S(P, Q; \ell, r)$ quantifies how similar their representations are at layer ℓ and resolution r. High similarity at one or more layers suggests an analogy that may be worth pursuing:

$$P \xrightarrow{\text{represent}} \{\phi(P, \ell, r)\} \xrightarrow{\text{similarity search}} Q \xrightarrow{\text{candidate invariant}} \xrightarrow{\text{transfer lemma}} \text{proof attempt.}$$

Crucially, similarity is a *discovery tool*, not a substitute for proof. The framework makes no claim that $S(P, Q; \ell, r) \approx 1$ implies that P and Q share a common solution: it only suggests that their representations are structurally close in layer ℓ at resolution r, and that a transfer of techniques may be fruitful.

12.3. The Role of the Visible/Hidden Decomposition

If at coarse resolution $R_r\phi(P, \ell) \approx R_r\phi(Q, \ell)$, then the visible component of the two problems looks similar. The hidden difference $D_{\text{hid}}(r)$ quantifies what remains unresolved at that resolution. In problem-solving terms, the hidden component corresponds to the structural details of P that the analogy does not capture, and that any transfer of a proof from Q to P must separately address:

control the hidden difference sufficiently to transfer the relevant property.

12.4. Impossibility as a Warning against False Analogy

Proposition 11.6 has a direct interpretation here: if the coarsening map π loses information (i.e. $\ker \pi \neq \{0\}$), then a proof at the coarser layer cannot be lifted to the finer layer without controlling the kernel component. In mathematical terms:

A proof at a coarse description level cannot generally be lifted to a finer level without additional co

This provides a formal explanation for a familiar phenomenon: analogies that work at an algebraic level may fail at a topological level, because the coarsening from topology to algebra typically has a non-trivial kernel (e.g. the lens spaces $L(7, 1)$ and $L(7, 2)$, which have identical homology groups $H_1 \cong \mathbb{Z}/7\mathbb{Z}$ but are not homotopy equivalent – so the algebraic coarsening cannot distinguish them while finer homotopy invariants can; or two graphs with the same degree sequence but different chromatic polynomials). The framework does not eliminate such failures; it *predicts* them and specifies what additional information would be needed to avoid them.

12.5. Research Hypothesis

> *Layer–resolution similarity analysis can serve as a systematic mechanism for discovering mathematically useful analogies, provided that any transfer of conclusions is separately justified by invariant-preserving maps or transfer lemmas, and that the hidden kernel component of the coarsening map is explicitly controlled.*

This hypothesis is in principle testable: it predicts that a similarity search over a structured corpus of problems and solutions, organised by $\mathcal{L}_{\text{math}}$ and $\mathcal{R}$, should surface known analogies (e.g. between combinatorics and algebra, or analysis and geometry) with higher precision than unstructured search, and should flag potential kernel failures before an analogy is pursued.

13. Proposed Application: Object Recognition under Variable Resolution

This section describes how the abstract framework of the preceding sections can be instantiated in the domain of object recognition in computer vision. The goal is not to claim superiority over existing systems, but to identify the specific structural elements that the framework adds to the standard pipeline and to state a testable hypothesis.

13.1. Background

Modern object detection systems such as YOLO [Terven et al., 2023], DETR [Carion et al., 2020], and Feature Pyramid Networks (FPN) [Lin et al., 2017] already exploit multiple scales and feature levels within a neural backbone. Speed–accuracy trade-offs in these systems have been studied extensively [Huang et al., 2017]. The present framework does not replace these architectures; rather, it proposes an additional structural layer on top of any backbone that makes the following concepts explicit and measurable:

(i) the distinction between *descriptive layer* and *resolution within a layer*;

(ii) the separation of total object-pair difference into *visible* and *hidden* components at each resolution;

(iii) cross-layer similarity bounds governed by the contraction ratio $\kappa(\ell_1, \ell_2)$.

13.2. Mapping the Framework to a Vision Pipeline

Let I denote an input image and $T^{(c)}$ a stored prototype for object class c. The framework maps naturally as follows:

$$I \longmapsto \phi(I,\ell) \longmapsto R_r^{(\ell)}\phi(I,\ell) \longmapsto S\big(I, T^{(c)}; \ell, r\big) = \Phi\Big(\big\|R_r^{(\ell)}\big(\phi(I,\ell) - \phi(T^{(c)},\ell)\big)\big\|_\ell\Big)\,.$$

Concrete layers can be assigned as follows:

$\ell_1 =$ edges/textures, $\ell_2 =$ local parts, $\ell_3 =$ object shape, $\ell_4 =$ semantic category, $\ell_5 =$ contex

Each layer admits a range of resolutions $r_1 \preceq r_2 \preceq \cdots$ corresponding to progressively finer feature representations.

Instead of returning a single scalar $P(\text{class})$, the system produces a *Layered Similarity Matrix* for each class:

$$\mathbf{S}^{(c)} = \big(S(I, T^{(c)}; \ell_i, r_j)\big)_{i,j},$$

which encodes, for each layer and resolution, how similar the input is to the class prototype.

13.3. Example: Ambiguity between Human and Cat

Suppose a low-resolution image contains a small figure. The system may compute, at coarse shape resolution:

$$S(\text{image}, \text{human}; \ell_3, r_1) = 0.82, \qquad S(\text{image}, \text{cat}; \ell_3, r_1) = 0.77.$$

The margin is insufficient for confident classification. The framework responds to this ambiguity by quantifying:

$$D_{\text{hid}}(r_1) = D_{\text{total}} - D_{\text{vis}}(r_1),$$

i.e. the fraction of distinguishing information not yet accessible at resolution r_1. This yields a principled statement: *there is insufficient accessible difference at the current resolution to distinguish between human and cat*, rather than a low-confidence softmax output that provides no structural explanation.

13.4. Four Proposed Advantages

(i) Adaptive resolution allocation. The system starts at r_1 (coarsest) and asks whether the accessible visible difference is sufficient to decide:

$$\text{if} \quad S_{(1)} - S_{(2)} > \tau \quad \Rightarrow \quad \text{stop and classify;} \quad \text{else} \quad r \to r' \succ r.$$

Here $S_{(1)}, S_{(2)}$ are the top two class similarities. This implements the resource-constrained distinguishability principle of Section 10: compute only as much resolution as needed for discrimination.

(ii) Layer-resolved explainability. Classification evidence can be broken down by layer:

$$\underbrace{S_{\ell_1,r} = 0.91}_{\text{texture}}, \quad \underbrace{S_{\ell_2,r} = 0.88}_{\text{parts}}, \quad \underbrace{S_{\ell_3,r} = 0.94}_{\text{shape}}, \quad \underbrace{S_{\ell_4,r} = 0.83}_{\text{semantic}}.$$

If the decisive layer is ℓ_3 (shape), the system can report: *classification is driven by the shape layer, with texture and semantic similarity providing corroborating evidence.*

(iii) Robustness under image degradation. When an image is blurred, compressed, or low-resolution, the visible difference $D_{\text{vis}}(r)$ can be computed explicitly. A confidence measure of the form $\text{conf} = F(S, D_{\text{vis}}/D_{\text{total}})$ ties confidence to the fraction of distinguishing information that is actually accessible, potentially improving calibration in degraded conditions.

(iv) Cross-layer consistency regularization. For learned admissible maps satisfying axiom (L6) exactly, Corollary 11.4 and Example 11.5 give $S_{\ell_1,r} \geq S^{\kappa}_{\ell_2,r}$. In practice, learned maps satisfy (L6) only approximately; Corollary 11.13 then gives the robust usable bound:

$$S_{\ell_1,r} \;\geq\; S^{\kappa+\delta^*_r}_{\ell_2,r},$$

where δ^*_r (Definition 11.16) is the minimal relative intertwining error, computable from the learned maps on a calibration set. A network that violates this bound at inference time signals representational instability and can serve as an anomaly or out-of-distribution detector.

13.5. Proposed Experimental Design

We do not build a new detector from scratch. The hypothesis-isolating experiment adds a *Layer–Resolution Similarity (LRS) module* on top of an existing backbone (e.g. ResNet-50 + FPN):

$$\underbrace{\text{backbone}}_{\text{fixed}} \;\to\; \{F_{\ell_1}, \ldots, F_{\ell_m}\} \;\to\; \underbrace{\text{LRS module}}_{\text{proposed addition}} \;\to\; \text{class + ambiguity output.}$$

Comparison: baseline detector versus baseline + LRS module. Evaluation metrics extend beyond mAP:

- mean Average Precision (mAP) at standard IoU thresholds;
- Expected Calibration Error (ECE) and reliability diagrams;
- out-of-distribution (OOD) ambiguity detection rate;
- FLOPs and latency under adaptive resolution stopping;
- performance at degraded resolutions: $100\%, 75\%, 50\%, 25\%, 12.5\%$ of original resolution, and under blur, noise, and JPEG compression.

The primary hypothesis is:

> *A backbone augmented with an LRS module achieves equal or better mAP at degraded resolution and improved calibration (lower ECE) relative to the same backbone without the module, at no more than* 10% *additional latency at full resolution.*

This hypothesis is falsifiable and does not require the LRS module to outperform YOLO or DETR on standard benchmarks at full resolution, where those systems have substantial engineering and training advantages.

Acknowledgements

This research was conducted independently. The author thanks the mathematical communities of measure theory, Hilbert space operator theory, information theory, and ultrametric geometry for the foundational work on which this framework builds.

A. Construct Assessment Table

The table below scores all 35 constructs on a 1–10 scale across four criteria: **F** = Formalization (mathematical precision), **V** = Mathematical validity of stated claims, **C** = Centrality to the theory, **N** = Novelty within this framework. An asterisk (*) denotes claims conditional on nested-projection hypotheses.

#	Construct	F	V	C	N	Status
1	$S(A, B; r)$	8.5	9.0	8.5	5.5	Solid baseline
2	$S(A, B; \ell, r)$	8.0	8.5	10	8.0	Core object
3	$R_r : X \to X_r$	9.5	9.5	9.0	3.0	Well-founded
4	$D_r = \|R_r(E_A - E_B)\|$	9.5	9.5	8.0	3.5	Strong
5	D_{total}	9.5	10	8.0	2.5	Reference quantity
6	$D_{\text{vis}}(\Lambda)$	9.5	10	9.0	5.0	Key construct
7	$D_{\text{hid}}(\Lambda)$	9.5	10	9.5	6.5	Key construct
8	$D_{\text{tot}} = D_{\text{vis}} + D_{\text{hid}}$	10	10	10	6.5	Central principle
9	Difference spectrum $\mathcal{D}_{AB}(k)$	9.5	9.5	7.5	3.5	Useful
10	$F_{AB}(\Lambda)$	9.5	9.5	8.5	4.5	Strong
11	$H_{AB} = 1 - F_{AB}$	9.5	10	8.0	4.0	Strong
12	(Ω, μ) distinction space	8.0	8.0	9.0	7.0	Needs axioms
13	$\delta_{AB}(\omega)$	8.0	8.5	8.5	5.5	Needs axioms
14	$V_r = D_r / M_r$	9.0	9.0	8.0	5.0	Non-monotone: note 6.3
15	Refinement invariance	9.5	9.5	9.0	6.5	Important
16	$\mathscr{P}_{AB}(r, t)$ surface	9.0	9.0	9.5	7.5	Candidate primary
17	$A \sim_r B$	10	10	8.5	3.0	Classical
18	Merge scale $m(A, B)$	9.5	9.0	8.0	5.0	Useful
19	Ultrametric	9.5	9.5*	7.0	2.5	Derived
20	Resolution as partition	10	10	8.5	3.0	Classical
21	Resolution lattice	9.5	9.5	7.0	2.0	Classical
22	Statistical resolution Δ_O	9.0	9.0	8.0	4.0	Useful
23	Δ_R resource-constrained	9.5	9.5	8.5	5.0	Strong
24	$R_\tau(A, B)$	9.0	9.0	7.5	5.0	Useful
25	$\mathfrak{R}(R)$	9.0	9.0	8.0	5.5	Promising
26	$\eta(R) = d\mathfrak{R}/dR$	8.0	8.5	6.5	4.5	Secondary
27	$S_{\ell,r}$ single-layer	8.5	8.5	9.5	7.0	Core
28	$S_{L',r}$ multi-layer	8.0	8.0	9.0	7.5	Needs $\mathcal{A}$
29	S_{all}	7.5	7.5	9.0	8.0	Needs ν
30	$\mathbf{S}_{A,B}(r)$ profile	9.0	9.0	10	8.5	Most promising
31	Similarity depth L_τ	7.0	7.5	7.0	6.0	Conditional
32	Layer breakpoint	7.0	7.5	7.0	6.0	Conditional
33	Layer identity $S_{\ell,r} = 1$	9.5	9.5	9.5	7.5	Strong

continued...

#	Construct	F	V	C	N	Status
34	All-layer identity	8.5	8.0	9.0	8.0	Needs $\mathcal{D}$
35	Naive feature counting	5.0	6.0	3.0	2.0	**Do not use**

* Under nested-projection hypotheses. F = Formalization (1–10), V = Mathematical validity, C = Centrality to theory, N = Novelty within this framework.

B. Worked Example: Exponential-Kernel Similarity Transport

This appendix demonstrates the complete transport theory (Theorems 11.1, 11.9, Proposition 11.6) in a fully concrete setting, computing S_{ℓ_1}, S_{ℓ_2}, κ, σ, and ε from explicit objects.

B.1. Setup

Let $\mathcal{F}_{\ell_2} = \mathbb{R}^3$ and $\mathcal{F}_{\ell_1} = \mathbb{R}^2$, both with the Euclidean norm. Define the layer-coarsening map as the orthogonal projection onto the first two coordinates:

$$\pi_{\ell_1,\ell_2}(x_1, x_2, x_3) = (x_1, x_2).$$

Let $\Phi(d) = e^{-d}$ (exponential kernel, so $\Phi(0) = 1$). Set the resolution operators to the identity: $R_r^{(\ell_2)} = I_{\mathbb{R}^3}$ and $R_r^{(\ell_1)} = I_{\mathbb{R}^2}$ (full resolution, no compression).

Take two objects A and B whose difference vector is

$$u_2 = \phi(A, \ell_2) - \phi(B, \ell_2) = (1,\ 0,\ 1).$$

B.2. Computing the Quantities

Layer contraction ratio κ.

$$\kappa(\ell_1, \ell_2) = \sup_{u \neq 0} \frac{\|\pi u\|_2}{\|u\|_2} = 1.$$

(The orthogonal projection achieves norm 1 on any vector with a unit-length component in the projected subspace; e.g. on $(1, 0, 0)$ the ratio is 1.)

Minimum gain σ. $(\ker \pi)^\perp = \{(x_1, x_2, 0)\}$, so

$$\sigma(\pi) = \inf_{\substack{u=(x_1,x_2,0) \\ u \neq 0}} \frac{\|(x_1, x_2)\|_2}{\|(x_1, x_2, 0)\|_2} = 1.$$

Distances.

$$D_r^{\ell_2}(A, B) = \|u_2\|_2 = \|(1, 0, 1)\|_2 = \sqrt{2},$$
$$u_1 = \pi(u_2) = (1, 0), \qquad D_r^{\ell_1}(A, B) = \|u_1\|_2 = 1.$$

Similarities.

$$S(A, B; \ell_2, r) = e^{-\sqrt{2}} \approx 0.243, \qquad S(A, B; \ell_1, r) = e^{-1} \approx 0.368.$$

B.3. Verifying the Bounds

Downward bound (Theorem 11.1).

$$S(A, B; \ell_1, r) \geq S(A, B; \ell_2, r)^{\kappa} = (e^{-\sqrt{2}})^1 = e^{-\sqrt{2}} \approx 0.243.$$

Observed: $0.368 \geq 0.243$. ✓

Upward impossibility (Proposition 11.6).

$$\ker \pi = \{(0, 0, x_3)\},$$

which is non-trivial ($x_3 = 1$ gives $\|\cdot\| = 1 > 0$, $\pi(\cdot) = 0$). Hence no finite global γ exists. ✓

Hidden kernel component. Decompose $v_2 = R_r^{(\ell_2)} u_2 = u_2 = (1, 0, 1)$:

$$v_{\|} = (1, 0, 0), \qquad v_{\ker} = (0, 0, 1), \qquad \varepsilon = \|v_{\ker}\|_2 = 1.$$

Conditional upward bound (Theorem 11.9).

$$D_r^{\ell_2}(A, B) \leq \frac{D_r^{\ell_1}(A, B)}{\sigma} + \varepsilon = \frac{1}{1} + 1 = 2.$$

Observed: $\sqrt{2} \approx 1.414 \leq 2$. ✓

In terms of similarity:

$$S(A, B; \ell_2, r) \geq \Phi\left(\frac{D_r^{\ell_1}}{\sigma} + \varepsilon\right) = e^{-2} \approx 0.135.$$

Observed: $0.243 \geq 0.135$. ✓

B.4. Interpretation

This example illustrates the complete asymmetric transport structure:

- Moving *down* from ℓ_2 to ℓ_1, we know $S_{\ell_1} \geq S_{\ell_2}^{\kappa} = S_{\ell_2}$, since $\kappa = 1$ introduces no contraction factor.
- Moving *up* from ℓ_1 to ℓ_2 globally is impossible: the third coordinate of u_2 is completely hidden at layer ℓ_1.
- Given the hidden kernel bound $\varepsilon = 1$ (we know that the invisible component is at most unit-length), we recover $S_{\ell_2} \geq e^{-2} \approx 0.135$, a non-trivial lower bound.

The gap between the actual similarity (0.243) and the bound (0.135) reflects the looseness when ε is set to its true value $\|v_{\ker}\| = 1$ rather than a tighter estimate. In this example, when $\varepsilon = 0$ for the specific pair (A, B), the difference vector u_2 has no kernel component; since also $\kappa = \sigma = 1$, the two-sided bounds collapse to $S(A, B; \ell_2, r) = S(A, B; \ell_1, r)$. This is a *pair-specific* statement: it says that this particular difference is fully visible at layer ℓ_1. It does not imply that the two layers are globally informationally equivalent; other difference vectors may have non-zero kernel components and will be compressed differently.

B.5. Second Example: Strict Separation $0 < \sigma < \kappa < 1$

The first example used $\kappa = \sigma = 1$, which did not exercise the separation between κ and σ. This example constructs a coarsening map achieving $0 < \sigma < \kappa < 1$ to illustrate the full asymmetry of the bounds.

Setup. Let $\mathcal{F}_{\ell_2} = \mathbb{R}^3$, $\mathcal{F}_{\ell_1} = \mathbb{R}^2$, $\Phi(d) = e^{-d}$, $R_r^{(\ell)} = I$ (full resolution). Define the coarsening map

$$\pi(x_1, x_2, x_3) = \left(\tfrac{3}{4}x_1,\ \tfrac{1}{2}x_2\right).$$

This simultaneously scales x_1 by $3/4$ and x_2 by $1/2$, while discarding x_3.

Computing κ and σ. The kernel is $\ker \pi = \{(0, 0, x_3)\}$ and $(\ker \pi)^\perp = \{(x_1, x_2, 0)\}$. On $(\ker \pi)^\perp$:

$$\frac{\|\pi(x_1, x_2, 0)\|}{\|(x_1, x_2, 0)\|} = \frac{\sqrt{\frac{9}{16}x_1^2 + \frac{1}{4}x_2^2}}{\sqrt{x_1^2 + x_2^2}}.$$

The supremum is achieved as $x_2 \to 0$, giving $\kappa = 3/4$. The infimum is achieved as $x_1 \to 0$, giving $\sigma = 1/2$. Hence:

$$\boxed{\sigma = \tfrac{1}{2} < \kappa = \tfrac{3}{4} < 1.}$$

Chosen difference vector. Take $u_2 = (0, 2, 1)$; $D_r^{\ell_2} = \sqrt{5} \approx 2.236$. Applying π: $u_1 = \pi(0, 2, 1) = (0, 1)$, $D_r^{\ell_1} = 1$. Similarities:

$$S_{\ell_2} = e^{-\sqrt{5}} \approx 0.108, \qquad S_{\ell_1} = e^{-1} \approx 0.368.$$

Decomposition. $v_2 = u_2 = (0, 2, 1)$; $v_\parallel = (0, 2, 0)$; $v_{\text{ker}} = (0, 0, 1)$; $\varepsilon = 1$.

Verifying the bounds. *Downward* (Theorem 11.1, exponential kernel, eq. (12)):

$$S_{\ell_1} \geq S_{\ell_2}^{\kappa} = e^{-\sqrt{5}\cdot 3/4} \approx e^{-1.677} \approx 0.187. \quad 0.368 \geq 0.187. \checkmark$$

Conditional upward (Theorem 11.9) with $\sigma = 1/2$, $\varepsilon = 1$:

$$D_{\ell_2} \leq \frac{D_{\ell_1}}{\sigma} + \varepsilon = \frac{1}{1/2} + 1 = 3. \quad \sqrt{5} \approx 2.236 \leq 3. \checkmark$$

In similarity:

$$S_{\ell_2} \geq e^{-3} \approx 0.050. \quad 0.108 \geq 0.050. \checkmark$$

Numerical summary and interpretation. The chain of inequalities is:

$$\underbrace{0.050}_{\text{upward bound}} \leq \underbrace{0.108}_{S_{\ell_2}} \leq \underbrace{0.187}_{\text{downward bound}} \leq \underbrace{0.368}_{S_{\ell_1}}.$$

The two bounds are controlled by different constants:

- *Downward bound* ($S_{\ell_1} \geq S_{\ell_2}^{3/4}$): controlled by $\kappa = 3/4$. A larger κ would tighten this bound toward S_{ℓ_2}; $\kappa = 1$ would give $S_{\ell_1} \geq S_{\ell_2}$.

- *Upward bound* ($S_{\ell_2} \geq e^{-3}$): controlled by $\sigma = 1/2$. A larger σ would allow more of D_{ℓ_1} to be attributed to the visible component, tightening the upward bound.

Since $\sigma < \kappa$, the two bounds are strictly governed by different quantities, demonstrating that κ *controls the downward direction and* σ *controls the upward direction* independently.

C. Proposed Application: Layer–Resolution Control of Large Language Models

This section proposes a third line of application, distinct from computer vision (Section 13) and mathematical analogy (Section 12). It is presented as a *research direction* with a falsifiable hypothesis; no empirical results are claimed.

C.1. Motivation

Transformer-based language models process an input token sequence through L successive layers, producing hidden states $h_1(x), \ldots, h_L(x)$ at each layer. Existing work exploits this structure in two ways. *Adaptive computation* methods reduce compute by skipping layers or exiting early when confidence is deemed sufficient [Schuster et al., 2022, Wee et al., 2025]. *Representation similarity* methods measure how much information is preserved or transformed across layers to guide compression or interpretability [Jiang et al., 2025].

Neither framework asks: *given the similarity between two candidate answers at layer* ℓ*, how much more computation is needed to achieve sufficient distinguishability*? The present framework provides the tools to answer this question.

C.2. Mapping the Framework to a Transformer

Let x be an input token, A and B two candidate answers, and p_c^ℓ a prototype for class/answer c at layer ℓ. The framework maps as follows:

$$h_\ell(x) \;\longrightarrow\; \phi(x,\ell) \;\longrightarrow\; R_r^{(\ell)}\phi(x,\ell) \;\longrightarrow\; S(A,B;\ell,r) = \Phi\big(\big\|R_r^{(\ell)}(h_\ell(A) - h_\ell(B))\big\|\big)\,,$$

where ℓ indexes a *sequence of abstract descriptive levels*, and r indexes resolution within each level.

Important caveat. Transformer hidden layers are not automatically admissible layer-coarsening maps in the sense of Definition 3.1: the map $h_{\ell+1} = f_\ell(h_\ell)$ is nonlinear and need not satisfy axioms (L3)–(L6). The proposed application therefore requires one of two approaches:

(A1) *Linearized coarsening*: choose a pair of representation spaces $(\mathcal{F}_{\ell_2}, \mathcal{F}_{\ell_1})$ and a direction of coarsening ($\ell_1 \preceq_{\mathcal{L}} \ell_2$ means ℓ_1 is coarser), then fit a linear map $\hat{\pi}_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ on a calibration set and verify contractivity (L5) and compatibility (L6) approximately. *Caution*: the forward Jacobian $J_{f_\ell} : \mathcal{F}_\ell \to \mathcal{F}_{\ell+1}$ maps *from* a coarser layer *to* a finer one and is not itself a coarsening map π; identifying the two would confuse computational depth with descriptive coarsening. The appropriate linear map must be constructed explicitly in the direction $\mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$.

(A2) *Learned admissible maps*: train explicit coarsening projections $\hat{\pi}_{\ell_1,\ell_2} : \mathcal{F}_{\ell_2} \to \mathcal{F}_{\ell_1}$ (e.g. linear probes) subject to a training objective that enforces *both* contractivity (L5) and approximate layer–resolution compatibility (17). This allows the robust bounds of Theorem 11.11 and Corollary 11.13 to apply, with δ_r^* (Definition 11.16) playing the role of the residual error. The exact bounds of Theorem 11.1 and Theorem 11.9 are recovered as the special case $\delta_r^* \to 0$.

The core research question is: *can admissible Layer–Resolution maps be learned from a real Transformer while preserving useful transport bounds?*

C.3. Four Proposed Uses

(i) Resolution-Aware Adaptive Computation. Current early-exit methods stop when model confidence (e.g. output entropy) exceeds a threshold. The proposed criterion is instead:

$$\text{exit at layer } \ell \quad \text{if} \quad D_{\text{vis}}(\ell, r) \geq D_{\text{required}},$$

i.e. when sufficient *visible* difference between the top candidates has accumulated. This links the stopping criterion directly to the distinguishability structure of the task rather than to output probability.

(ii) Hidden-Difference Hallucination Score. Define the *Layer–Resolution Ambiguity Score* (LRAS):

$$\text{LRAS}(x; \ell, r) = P_{\max} \cdot \max\left\{0,\ 1 - \frac{D_{\text{vis}}(\ell, r)}{D_{\text{required}}}\right\},$$

where $P_{\max}$ is the model's output confidence and D_{required} is a task-specific distinguishability threshold. The $\max\{0, \cdot\}$ clipping ensures LRAS $\in [0, 1]$: when $D_{\text{vis}} \geq D_{\text{required}}$, visible difference already exceeds the required amount and LRAS is zero (no ambiguity detected). High LRAS signals *high confidence despite low internal distinguishability*: a potential hallucination or failure mode. This complements entropy-based uncertainty measures [Dai et al., 2026] by locating the layer and resolution at which the ambiguity originates.

(iii) Cross-Layer Consistency Regularizer. For learned admissible maps satisfying (L6) exactly, Corollary 11.4 gives the ideal bound $S_{\ell_1} \geq S_{\ell_2}^{\kappa}$. In practice (approach (A1) or (A2)), the learned map satisfies (L6) only approximately; the operationally correct bound from Corollary 11.13 is:

$$S_{\ell_1} \geq S_{\ell_2}^{\kappa + \delta_r^*},$$

where δ_r^* (Definition 11.16) is computable from the learned maps on a calibration set. Violations of this bound at inference time – where the observed S_{ℓ_1} is strictly below $S_{\ell_2}^{\kappa+\delta_r^*}$ – signal representational instability; this can serve as an *anomaly detector* for adversarial prompts, out-of-distribution inputs, or corrupted representations, without requiring a separate classifier.

(iv) Information-Loss-Bounded Compression. When compressing the key–value cache or pruning tokens, Proposition 11.6 gives a principled criterion:

$$\text{compress only if } \varepsilon = \|P_{\ker \pi} R_r^{(\ell)} u_{AB}\| \leq \tau.$$

That is, remove a representation only when the hidden kernel component of the relevant difference is provably below a tolerance τ, rather than using heuristic pruning criteria [Hooper et al., 2025].

C.4. Proposed Experimental Design

The hypothesis to test is:

> *A Transformer augmented with a Layer–Resolution Similarity Controller achieves equal or better accuracy at degraded input conditions and lower LRAS-flagged hallucination rate, at no more than* 10% *additional latency at full resolution, relative to the same Transformer with standard early-exit or without a controller.*

The proposed experimental pipeline:

(i) Record hidden states $\{h_\ell(x)\}$ from an open-weight Transformer (no modification to the base model).

(ii) Construct admissible coarsening maps $\hat{\pi}_{\ell_i,\ell_j}$ (via linearization (A1) or constrained learning (A2)), and compute: $\kappa_{ij} = \|\hat{\pi}_{ij}\|$ (largest singular value), σ_{ij} (smallest non-zero singular value on $(\ker \hat{\pi})^\perp$), and $\delta^*_{r,ij}$ (largest singular value of $(P\hat{\pi} - \hat{\pi}R)R^\dagger$ by Proposition 11.17).

(iii) Deploy a lightweight LR Similarity Controller on top of the frozen Transformer.

(iv) Compare: full-depth baseline vs. standard early-exit vs. LR-controlled exit.

(v) Evaluate: accuracy, LRAS, Expected Calibration Error, FLOPs, latency, and false-confident error rate.

References

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953. doi: 10.1214/aoms/1177729032.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. *arXiv preprint*, arXiv:2005.12872, 2020. URL https://arxiv.org/abs/2005.12872.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006. ISBN 978-0471241959.

Zeyu Dai, Lei Shen, Qiang Chen, Bo Wang, and Jialin He. Multi-scale adaptive semantic entropy framework: Hierarchical detection and early warning mechanism for large language model hallucinations. *Neurocomputing*, 2026. URL https://www.sciencedirect.com/science/article/pii/S0925231226013986.

Paul R. Halmos. *A Hilbert Space Problem Book*. Springer, 2nd edition, 1982. doi: 10.1007/978-1-4684-9330-6.

Coleman Hooper, Sehoon Kim, Hasan Anil Anil, Mostafa Elhoushi, Yejin Kim, Sheng Shen, Michael W. Mahoney, Kurt Keutzer, and Amir Gholami. QuickSilver: Speeding up LLM inference through dynamic token halting, KV skipping, contextual token fusion, and adaptive Matryoshka quantization. *arXiv preprint arXiv:2506.22396*, 2025. URL https://arxiv.org/abs/2506.22396.

Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7310–7311, 2017. doi: 10.1109/CVPR.2017.351.

Francesco Insulla, Shuo Huang, and Lorenzo Rosasco. Towards a learning theory of representation alignment. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/5d4f5a2de6320641566be8722d5f78dc-Abstract-Conference.html.

Kazuo Ishii, Bishnu Prasad Gautam, Jieling Wu, and Javaid Saher. Platonic projection structures: Operator-induced observability in representation learning. *arXiv preprint arXiv:2607.05175*, 2026. URL https://arxiv.org/abs/2607.05175.

Jiachen Jiang, Jinxin Zhou, and Zhihui Zhu. Tracing representation progression: Analyzing and enhancing layer-wise similarity. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/03d113a060c0ac93a5859517a0f07271-Abstract-Conference.html.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, 2nd (reprint) edition, 1995. ISBN 978-3540586616. doi: 10.1007/978-3-642-66282-9.

Max Klabunde, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Similarity of neural network models: A survey of functional and representational measures. *ACM Computing Surveys*, 57(9), 2025. doi: 10.1145/3728458.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. *arXiv preprint arXiv:1905.00414*, 2019. URL https://arxiv.org/abs/1905.00414.

Brian Kulis. Metric learning: A survey. *Foundations and Trends in Machine Learning*, 5(4):287–364, 2013. doi: 10.1561/2200000019.

Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2117–2125, 2017. doi: 10.1109/CVPR.2017.106.

Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994. ISBN 978-0792394181. doi: 10.1007/978-1-4757-6465-9.

R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Reviews of Modern Physics*, 58(3):765–788, 1986. doi: 10.1103/RevModPhys.58.765.

Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics, Vol. I: Functional Analysis*. Academic Press, 1980. ISBN 978-0125850506.

Tal Schuster, Adam Fisch, Jai Gupta, Mostafa Dehghani, Dara Bahri, Vinh Q. Tran, Yi Tay, and Donald Metzler. Confident adaptive language modeling. *Advances in Neural Information Processing Systems*, 35:17456–17472, 2022. URL https://arxiv.org/abs/2207.07061.

Shaan Shah and Meenakshi Khosla. Representational alignment across model layers and brain regions with multi-level optimal transport. In *Proceedings of the 14th International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8611049567e7339a88fc68a06100fac1-Abstract-Conference.html.

Juan Terven, Diana-Margarita Cordova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4):1680–1716, 2023. doi: 10.3390/make5040083.

Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977. doi: 10.1037/0033-295X.84.4.327.

Justin Wee et al. Prompt-based depth pruning of large language models. In *Proceedings of Machine Learning Research*, volume 267, 2025. URL https://proceedings.mlr.press/v267/wee25a.html.

Andrew P. Witkin. Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1019–1022, 1983.