mathematics

MDPI

Article

Finite AMN-Inspired Geometric Regularization for Neural Metric Learning

Alberto Muñoz

Department of Statistics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain; alberto.munoz@uc3m.es

Abstract

Neural metric learning is often assessed by retrieval accuracy, but a learned dissimilarity can rank examples well while failing to have norm-like algebraic structure. This paper studies a precise finite question: within a Euclidean-anchored residual family of neural dissimilarities, can one reduce sampled defects of homogeneity, subadditivity, and dyadic reconstruction on latent differences without destroying retrieval performance? The construction is inspired by asymptotically metrically normable (AMN) vector spaces, but its claims are finite, sampled, and latent: it does not prove global AMN rigidity or certify a metric on the input space. The framework is motivated by the observation that many learned similarities have the form $K = \exp(-E/\tau)$ and therefore encode an unbounded distance-like quantity or squared distance-like quantity behind a bounded affinity. The AMN-relevant object is this cost, not the bounded kernel value. We formalize bounded-perturbation stability of the large-scale specific energy $E(nv, 0)/n$, the conversion of subadditivity into multiplicative affinity consistency, and the quotient interpretation in which directions of zero large-scale cost are collapsed. The mathematical development then introduces finite dyadic diagnostics, learned-gauge and convex-unit-ball interpretations, finite norm-envelope witnesses, dyadic stability bounds, and refinement towers of witness norms. The empirical part reports full official Fashion-MNIST experiments with supervised-contrastive and proxy-anchor-style Euclidean baselines, post hoc audits for shrinkage, residual flexibility, off-training scales, and latent extrapolation, and a ten-seed full-query/full-gallery UCI Human Activity Recognition benchmark. The results show that Euclidean objectives can be stronger for Recall@1, whereas AMN-inspired residual regularization substantially reduces finite norm-like defects inside the residual family. The contribution is therefore a finite diagnostic and regularization framework for learned latent dissimilarities, not a state-of-the-art retrieval objective.

Keywords: neural metric learning; neural dissimilarities; asymptotically metrically normable spaces; AMN-inspired regularization; geometric regularization; norm reconstruction; finite diagnostics; kernel-derived distances; quotient latent geometry; Fashion-MNIST; UCI HAR; sensor data

MSC: 68T07; 46B20; 54E35; 68T09; 68T10

Check for updates

Academic Editor: Jay M. Jahangiri

Received: 5 June 2026
Revised: 29 June 2026
Accepted: 3 July 2026
Published: 6 July 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Metric-learning algorithms are commonly judged by neighbor ranking, same-class retrieval, or embedding utility. Contrastive losses, triplet losses, and supervised contrastive objectives are effective for these tasks [1–3]. However, good retrieval performance does

not by itself imply that the learned dissimilarity has the algebraic structure of a norm. Pairwise and triplet objectives constrain relative scores on sampled examples; they do not necessarily enforce homogeneity, subadditivity, or stable norm reconstruction [4,5]. A neural dissimilarity may therefore be useful for retrieval while still having scale defects or non-subadditive behavior.

This paper focuses on the following finite claim. Consider a residual neural dissimilarity evaluated on latent differences and already anchored to the Euclidean norm. We ask whether AMN-inspired finite penalties can make this dissimilarity more norm-like on sampled latent scales, while preserving most of its retrieval performance. The question is intentionally narrower than global metric certification. All diagnostics and guarantees in the paper concern finite samples of latent vectors and finite-scale sets; they do not prove that a trained network induces a metric or a norm on the original input space.

The motivation comes from Mu noz-García's theory of asymptotically metrically normable (AMN) vector spaces [6]. In that theory, norm-like geometry is recovered from the large-scale behavior of a metric vector space. The present paper does not transfer AMN rigidity theorems to neural networks. Instead, it extracts an operational principle: if a learned dissimilarity is approximately homogeneous, approximately subadditive, and stable under a finite slope reconstruction, then it can be audited and regularized as a finite norm-like object.

A useful representation-learning example comes from exponential similarities. Suppose latent objects lie in a vector space E and a learned similarity has the form

$$K(x,y) = \exp\{-D(x,y)/\tau\} \tag{1}$$

for a Laplacian or exponential model, or $K(x,y) = \exp\{-D(x,y)^2/(2\ell^2)\}$ for a Gaussian model. In the first case, $D(x,y) = -\tau \log K(x,y)$ is already distance-like; in the second case, the logarithm gives $D(x,y)^2 = -2\ell^2 \log K(x,y)$ and the square root is the distance-like quantity. If, up to bounded local perturbations,

$$D(x,y) = \|A(x-y)\|_2 + \varepsilon(x,y), \qquad |\varepsilon(x,y)| \le B, \tag{2}$$

then

$$\lim_{n\to\infty} \frac{D(nv,0)}{n} = \|Av\|_2, \tag{3}$$

while the bounded directions are precisely $\ker A$. Thus, the effective large-scale geometry lives on the quotient $E/\ker A$: directions in $\ker A$ are invariances, redundant coordinates, noise, or collapse, whereas the quotient carries the stable semantic norm. The finite framework developed below is a sampled analog of this quotient picture.

Let X be the input space, let $f_\psi : X \to \mathbb{R}^m$ be an encoder and write $z_i = f_\psi(x_i)$. The residual head is evaluated on differences $v = z_i - z_j$ and has the form

$$\rho_\theta(v) = \|v\|_2\big(1 + a\tanh(g_\theta(v))\big) + \delta, \tag{4}$$

where $g_\theta : \mathbb{R}^m \to \mathbb{R}$ is a scalar neural head, $a \in [0,1)$ bounds the residual deformation and $\delta > 0$ is a numerical stabilizer. The dissimilarity used in diagnostics is the symmetrized version

$$d_\theta(v) = \frac{\rho_\theta(v) + \rho_\theta(-v)}{2}. \tag{5}$$

For a finite dyadic depth L, the associated slope estimator is

$$\widehat{N}_L(v) = \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k} d_\theta(2^k v). \tag{6}$$

This estimator is a diagnostic, not a certified norm. It is chosen because dyadic scaling gives an explicit telescoping identity and a computationally simple finite analog of large-scale AMN slopes; off-grid audits later check whether the observed behavior is not restricted to the training scales.

The main contribution is a finite diagnostic and regularization framework for learned latent dissimilarities. Its theoretical component explains the diagnostics through dyadic slope stability, learned gauges, convexity of finite unit balls, finite norm-envelope witnesses, and refinement towers of witness norms. Its experimental component tests whether the resulting penalties reduce finite defects in a controlled residual family, while modern Euclidean metric-learning baselines provide retrieval context. The full-split Fashion-MNIST experiments show that supervised contrastive Euclidean embeddings can achieve higher Recall@1, whereas AMN-inspired terms reduce finite norm-like defects inside the residual family. UCI HAR provides a non-visual secondary check with the same qualitative residual-family signal.

The rest of the paper is organized as follows. Section 2 recalls the AMN motivation, kernel-energy interpretation, and related metric-learning work. Section 3 defines the residual dissimilarity, the finite diagnostics, the norm-envelope and refinement constructions, the training objective, and the experimental protocol. Section 4 reports Fashion-MNIST, UCI HAR, and supporting audits. Section 5 interprets the results, including the role of Euclidean anchoring and the limitations of the diagnostics. Section 6 summarizes the contribution and future directions.

2. Background and Related Work

2.1. AMN Vector Spaces and Finite Translation

Mu noz-García's AMN theory studies metric vector spaces whose large-scale geometry is asymptotically normable [6]. For an exact norm N, the identities

$$N(\lambda v) = |\lambda| N(v), \qquad N(v+w) \le N(v) + N(w) \tag{7}$$

encode scaling and subadditivity. AMN theory replaces such identities with asymptotic controls and reconstructs norm-like behavior from large-scale slopes such as

$$\lim_{n\to\infty} \frac{1}{n} d(nx, 0). \tag{8}$$

The present paper uses this theory as motivation, not as a direct theorem about trained neural networks. We work with finite latent differences $v = z_i - z_j \in \mathbb{R}^m$ and replace infinite large-scale limits by the finite dyadic estimator Equation (6). Table 1 separates the AMN ingredient, the finite neural counterpart, and the corresponding limitation. This separation is important: the rigorous statements below are finite stability, envelope, and refinement results; the AMN theory supplies the guiding norm-reconstruction viewpoint.

Table 1. Translation from asymptotic AMN theory to the finite diagnostics used here. The table records which part is rigorous in the present finite model and what is not claimed.

Asymptotic AMN Object	Finite Neural Analog Used Here	What Is Not Claimed
Translation-invariant vector metric $d(x,y) = d(x-y,0)$	Parametrize the learned dissimilarity by latent differences $v = f_\psi(x_i) - f_\psi(x_j)$	Translation invariance is built into the latent parametrization; it is not certified for the original input space.
Asymptotic scaling and slope reconstruction, e.g., $n^{-1}d(nx,0)$	Finite dyadic slopes $2^{-k}d_\theta(2^k v)$ and the estimator $\widehat{N}_L(v)$ for $k = 0, \dots, L$	No convergence as $L \to \infty$ or control over all scalar factors is proved.

Table 1. *Cont.*

Asymptotic AMN Object	Finite Neural Analog Used Here	What Is Not Claimed
Controlled sums at large scale	Minibatch subadditivity defect D_{sub} on sampled latent pairs (v, w)	No global triangle inequality is guaranteed; $v + w$ need not be an observed data difference.
Existence of a norm asymptotically equivalent to the metric	Empirical reduction of homogeneity, subadditivity and reconstruction defects	No normability theorem or Lipschitz equivalence to a true norm is established.
Distance-like quantity behind an exponential kernel	Apply finite diagnostics to $D = -\tau \log K$ for Laplacian forms or to $D = \sqrt{-2\ell^2 \log K}$ for Gaussian forms; the raw negative log is squared-distance-like in the Gaussian case	Positive semidefiniteness of a reconstructed kernel requires additional assumptions beyond norm-like distance.
Rigidity under coarse perturbations	Stability of the finite estimator under controlled sampled defects	No AMN rigidity result is transferred to neural networks.

2.2. Kernel-Derived Distances and Quotient Latent Geometry

Exponential similarities often hide an unbounded cost behind a bounded affinity. Applying AMN scaling directly to a bounded normalized RKHS distance,

$$d_K(x,y)^2 = K(x,x) + K(y,y) - 2K(x,y), \tag{9}$$

would be uninformative when K is bounded, because all large-scale directions become bounded. The AMN-relevant object is instead the exponent. For a Laplacian or exponential form $K(x,y) = \exp(-D(x,y)/\tau)$, one recovers the distance-like quantity $D(x,y) = -\tau \log K(x,y)$. For a Gaussian form

$$K(x,y) = \exp\left(-\frac{\|A(x-y)\|_2^2}{2\ell^2}\right), \tag{10}$$

$-2\ell^2 \log K(x,y)$ is squared-distance-like and

$$D(x,y) = \sqrt{-2\ell^2 \log K(x,y)} = \|A(x-y)\|_2 \tag{11}$$

is the corresponding distance-like quantity. If a learned distance-like quantity satisfies $D(x,y) = \|A(x-y)\|_2 + O(1)$, then its AMN slope recovers $v \mapsto \|Av\|_2$ and the directions invisible at large scale are $\ker A$. Thus, $E/\ker A$ is the effective quotient latent space. This quotient interpretation is useful only under a stable large-scale model; it should not be applied to an arbitrary bounded kernel value.

The conversion back from a recovered norm-like distance to a positive-semidefinite kernel is a separate issue. Hilbertian or conditionally negative type assumptions can justify kernels such as $\exp(-\nu([x-y])/\tau)$ or $\exp(-\nu([x-y])^2/(2\sigma^2))$ [7,8], but positive semidefiniteness is not automatic for a general norm-like distance. We therefore distinguish a stable distance-like energy from any downstream PSD kernel.

2.3. Specific Energy, Affinity Consistency and Information Viewpoint

The notation $K = \exp(-E/\tau)$ is also useful as a compact energy analogy. Here, E is a displacement cost, not a conserved physical energy. The stable quantity is the specific large-scale slope

$$\nu(v) = \lim_{n\to\infty} \frac{E(nv, 0)}{n}, \tag{12}$$

when it exists. Bounded perturbations do not change this slope: $E' = E + O(1)$ gives the same asymptotic specific energy. In addition, if E is subadditive, then the exponential affinity is multiplicatively path-consistent:

$$E(x,z) \le E(x,y) + E(y,z) \quad \Longrightarrow \quad K_E(x,z) \ge K_E(x,y)K_E(y,z). \tag{13}$$

This is the kernel-language counterpart of the subadditivity defect used later.

The entropy connection is only pointwise. A raw value $-\log K(x,y)$ is a logarithmic cost, not an entropy. After normalizing affinities into a neighborhood distribution

$$p(y \mid x) = \frac{\exp(-E(x,y)/\tau)}{Z(x)}, \qquad Z(x) = \sum_z \exp(-E(x,z)/\tau), \tag{14}$$

the quantity $-\log p(y \mid x) = E(x,y)/\tau + \log Z(x)$ is a surprisal term and its expectation is a Shannon entropy [9,10]. Thus, the AMN-inspired diagnostics act on geometric cost, while entropy describes uncertainty of the induced neighbor distribution.

2.4. Scope of the AMN Connection

The AMN terminology is used in a restricted and operational sense. Translation invariance is built into the difference parametrization $v = z_i - z_j$ rather than learned as an input-space property. Vectors such as $v + w$ or $4v$ are ambient latent-vector operations and may not be observed data differences. Small sampled defects imply controlled behavior on the sampled latent distribution and finite scales; they do not certify a global metric, a norm on the input space, or a finite necessary-and-sufficient AMN criterion.

2.5. Metric Learning and Geometry of Learned Dissimilarities

Classical metric learning includes large-margin nearest-neighbor learning [11], Siamese and contrastive approaches [1,12], triplet-based deep metric learning [3], and supervised contrastive learning [2]. More recent pair-weighting and proxy-based objectives, including Circle Loss and Proxy Anchor, improve optimization and retrieval behavior [13,14]; surveys emphasize loss design, sampling, and evaluation protocol [4,5,15–17]. These objectives optimize task relations among samples, but they do not necessarily enforce homogeneity, subadditivity, or norm reconstruction for a learned dissimilarity.

The baselines answer two questions. Euclidean triplet, supervised-contrastive, and proxy-anchor-style baselines contextualize retrieval performance; they are norm-like by construction and are not expected to have large finite defects. The residual task-only model is the main geometric comparator, because it has the same residual capacity without AMN-inspired penalties. The non-AMN controls test whether the effect is generic regularization. The experiments should therefore be read as a controlled study of finite geometric regularization, complemented by standard retrieval baselines, rather than as a claim of state-of-the-art retrieval accuracy.

3. Materials and Methods

All computational experiments and analyses were run with Python 3.12.11. The principal software packages were PyTorch 2.7.1+cu118, NumPy 2.0.2, pandas 2.3.2, scikit-learn 1.6.1, and Matplotlib 3.10.6. No laboratory instruments, reagents, or commercial measurement agents were used.

3.1. Latent Residual Dissimilarity

Let (x_i, y_i) denote an input-label pair, where $x_i \in X$ is an observation and y_i is its class label. The observation may be an image or a tabular feature vector. The en-

coder $f_\psi : X \to \mathbb{R}^m$ maps inputs to latent vectors $z_i = f_\psi(x_i)$, and all learned dissimilarities are evaluated on latent differences $v = z_i - z_j$. The residual function ρ_θ is defined in Equation (4). In the final experiments, $m = 32$, $a = 0.65$ and $\delta = 10^{-7}$. Since $\tanh(g_\theta(v)) \in [-1, 1]$ and $a < 1$, the multiplicative factor in Equation (4) lies between $1 - a$ and $1 + a$, so the Euclidean factor prevents negative values and anchors the model to a norm-like reference. In particular, apart from the small stabilizer, the residual family is bi-bounded by the Euclidean norm: $(1-a)\|v\|_2 \le \rho_\theta(v) - \delta \le (1+a)\|v\|_2$. For the final value $a = 0.65$, this is the interval $[0.35\|v\|_2, 1.65\|v\|_2]$. Therefore, the experiment asks whether finite AMN-inspired penalties reduce residual geometric defects inside this Euclidean-anchored family, not whether they discover normability from an arbitrary dissimilarity.

The raw residual head g_θ need not be even in v. For this reason, the function used for retrieval, training diagnostics, and AMN-inspired losses is the symmetrized dissimilarity d_θ defined in Equation (5). By construction, $d_\theta(v) = d_\theta(-v)$. Strictly speaking, because of the stabilizer and because triangle inequalities are only regularized empirically, d_θ is used as a non-negative symmetric dissimilarity rather than as a certified metric. The additive stabilizer is retained in numerical denominators to avoid division by values close to zero.

Remark 1 (Latent-vector nature of the regularizer)**.** *The finite AMN-inspired terms are applied after encoding, not in the raw input space. Translation invariance is therefore imposed by the difference parametrization $v = z_i - z_j$. The vectors $v + w$ and λv used in the losses are latent-vector operations and need not be observed pair differences. The regularizer should therefore be interpreted as controlling the ambient latent extension of the learned dissimilarity on sampled scales.*

Figure 1 summarizes the latent-difference workflow, and Table 2 lists the main objects used in the definitions and diagnostics.

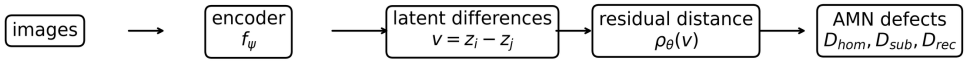


Figure 1. Pipeline. Inputs are encoded into latent vectors. The residual dissimilarity is evaluated on differences $v = z_i - z_j$, and finite AMN-inspired diagnostics measure homogeneity, subadditivity, and reconstruction defects on sampled latent-vector operations.

Table 2. Main notation. The table separates trained objects from derived diagnostics and records the status of each quantity.

Symbol	Domain/Construction	Trained or Derived	Meaning and Status
f_ψ	$X \to \mathbb{R}^m$	trained	Encoder producing latent vectors.
z_i	$f_\psi(x_i) \in \mathbb{R}^m$	derived	Latent representation of input x_i.
v	$z_i - z_j \in \mathbb{R}^m$	derived	Latent difference; regularization acts in this vector space.
ρ_θ	$\mathbb{R}^m \to \mathbb{R}_+$	trained residual head	Raw Euclidean-anchored dissimilarity; not a certified metric.
d_θ	$(\rho_\theta(v) + \rho_\theta(-v))/2$	derived	Even dissimilarity used for retrieval and diagnostics.
$\bar{\rho}$	$(\rho(v) + \rho(-v))/2$	derived	Generic symmetrization used in finite-stability statements.
$\widehat{N}_L$	finite average of $2^{-k}\rho(2^k v)$	derived	Dyadic norm-like estimator; diagnostic, not a certified norm.
$D_{\mathrm{hom}}, D_{\mathrm{sub}}, D_{\mathrm{rec}}$	expectations over sampled latent differences	reported/trained	Finite defects of scaling, subadditivity and dyadic reconstruction.
p_S	gauge of $\mathrm{conv}\{\pm v/\rho(v) : v \in S\}$	derived	Finite witness norm on $\mathrm{span}(S)$ when S spans.
Γ_S	$\sup_{u \in S}(1 - p_S(u)/\rho(u))_+$	derived	Finite norm-envelope gap on sampled differences.
E, D	exponent cost behind $K = \exp(-E/\tau)$	derived/modelled	Distance-like energy; for Gaussian kernels $D = \sqrt{-2\ell^2 \log K}$.

3.2. Finite AMN-Inspired Defects

All finite diagnostic quantities are computed from minibatch samples of latent differences. Let $\mathcal{V}_B$ denote the empirical distribution that samples latent pair differences $z_i - z_j$ from the vectors in minibatch B. The implementation samples the two indices independently with replacement, so exact self-pairs may occur with probability $1/|B|$; the numerical stabilizer in the denominators prevents division by zero in those rare cases. When several variables are written under an expectation, they are sampled independently unless stated otherwise. Let $\Lambda = \{1/2, 2, 4\}$ be the set of scaling factors used in the trained losses and primary reported diagnostics, let $\varepsilon = 10^{-7}$ be a numerical stabilizer, and let $[t]_+ = \max\{t, 0\}$. For a non-negative even dissimilarity ρ on latent differences, define the finite dyadic estimator by Equation (6), replacing d_θ by ρ when necessary. In the experiments, $\rho = d_\theta$ for diagnostics and $L = 4$ for the principal Fashion-MNIST runs.

The homogeneity defect measures whether scaling v by λ scales the value by $|\lambda|$:

$$D_{\mathrm{hom}}(\rho) = \mathbb{E}_{\substack{v\sim\mathcal{V}_B \\ \lambda\sim\mathrm{Unif}(\Lambda)}} \left| \frac{\rho(\lambda v)}{|\lambda|\rho(v) + \varepsilon} - 1 \right|. \tag{15}$$

The subadditivity defect measures normalized violations of the triangle inequality in latent-difference form:

$$D_{\mathrm{sub}}(\rho) = \mathbb{E}_{\substack{v,w\sim\mathcal{V}_B \\ \text{independent}}} \frac{[\rho(v+w) - \rho(v) - \rho(w)]_+}{\rho(v) + \rho(w) + \varepsilon}. \tag{16}$$

The reconstruction defect compares the learned dissimilarity with its finite dyadic estimator:

$$D_{\mathrm{rec}}(\rho) = \mathbb{E}_{v\sim\mathcal{V}_B} \frac{|\rho(v) - \widehat{N}_L(v)|}{\rho(v) + \varepsilon}. \tag{17}$$

The reported diagnostics use the absolute normalized quantities in Equations (15)–(17). The differentiable training losses use the corresponding squared normalized quantities, as defined in Definition 2.

The dyadic estimator is used because repeated doubling gives a telescoping identity in Proposition 1, is cheap to evaluate during training, and matches the large-scale-slope intuition behind AMN reconstruction. The choice is not unique. Other finite-scale grids could be used; this is why the experiments include off-training-scale audits at non-dyadic and intermediate factors. The finite results below therefore justify the estimator on its chosen scale set rather than proving the uniqueness of dyadic scaling. Table 3 summarizes the reporting and regularization roles of the three defects.

Table 3. Roles of the finite diagnostic quantities. The same quantities are used for reporting and, in squared form, for regularization, but none of them is by itself a global normability certificate.

Quantity	Operational Role	What It Does Not Certify
D_{hom}	Finite scaling diagnostic and loss term	Global homogeneity on all $\mathbb{R}^m$
D_{sub}	Finite subadditivity diagnostic and loss term	Triangle inequality for all triples in X
D_{rec}	Stability of d_θ under $\widehat{N}_L$	Existence of a limiting norm as $L \to \infty$
$\widehat{N}_L$	Finite dyadic slope estimator	A certified norm unless additional hypotheses hold

3.3. Finite-Stability Statement

The following proposition is the finite mathematical bridge between AMN intuition and the implemented estimator. It is a stability statement for Equation (6), not a charac-

terization of asymptotic AMN normability. The statement is written for a symmetrized dissimilarity, which matches Equation (5).

Proposition 1 (Finite stability of dyadic AMN-inspired reconstruction)**.** *Let $\rho : \mathbb{R}^m \to \mathbb{R}_+$ be a non-negative function and let $\bar{\rho}(v) = (\rho(v) + \rho(-v))/2$. Define*

$$a_k(v) = 2^{-k}\bar{\rho}(2^k v), \qquad \widehat{N}_L(v) = \frac{1}{L+1}\sum_{k=0}^{L} a_k(v). \tag{18}$$

Then:

(i) The dyadic homogeneity boundary identity is

$$\widehat{N}_L(2v) - 2\widehat{N}_L(v) = \frac{2}{L+1}\big(a_{L+1}(v) - a_0(v)\big). \tag{19}$$

Thus, if normalized slopes $a_k(v)$ vary little over the observed dyadic range, $\widehat{N}_L$ is approximately homogeneous.

(ii) If, for every $k = 0, \ldots, L$,

$$\bar{\rho}(2^k(v+w)) \le \bar{\rho}(2^k v) + \bar{\rho}(2^k w) + \varepsilon_s\big(1 + \bar{\rho}(2^k v) + \bar{\rho}(2^k w)\big), \tag{20}$$

then

$$\widehat{N}_L(v+w) \le (1+\varepsilon_s)\big(\widehat{N}_L(v) + \widehat{N}_L(w)\big) + \varepsilon_s c_L, \qquad c_L = \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k} \le \frac{2}{L+1}. \tag{21}$$

(iii) The reconstruction error satisfies

$$|\widehat{N}_L(v) - \bar{\rho}(v)| \le \frac{1}{L+1}\sum_{k=0}^{L} |a_k(v) - a_0(v)|. \tag{22}$$

The proof is given in Appendix A. The proposition formalizes the diagnostic intuition: homogeneity controls variation of normalized slopes, subadditivity controls finite triangle behavior, and both improve the stability of the reconstructed estimator. Its practical significance is finite consistency. It gives a reason for monitoring and regularizing the three defects together; it does not prove that small empirical losses certify a global norm or metric.

The next statement makes the link between dyadic homogeneity and reconstruction error more explicit. It is stated along one ray and uses log-slopes to avoid asymmetric ratio bounds.

Proposition 2 (Relative dyadic slope control)**.** *Let $v \neq 0$ and suppose that $a_k(v) > 0$ for $k = 0, \ldots, L$. If there is $\eta \ge 0$ such that*

$$|\log a_{k+1}(v) - \log a_k(v)| \le \eta, \qquad k = 0, \ldots, L-1, \tag{23}$$

then, for every $k = 0, \ldots, L$,

$$e^{-k\eta} a_0(v) \le a_k(v) \le e^{k\eta} a_0(v), \tag{24}$$

and the relative reconstruction error satisfies

$$\frac{|\widehat{N}_L(v) - \bar{\rho}(v)|}{\bar{\rho}(v)} \leq \frac{1}{L+1} \sum_{k=0}^{L} \left(e^{k\eta} - 1 \right). \tag{25}$$

For small η, the right-hand side is $\frac{L}{2}\eta + O(L^2\eta^2)$. Thus, finite dyadic homogeneity gives a quantitative finite control of the dyadic reconstruction error along the same ray.

A second geometric interpretation is useful because the residual family has the multiplicative form $\|v\|_2 q_\theta(v)$, up to the small stabilizer. This places the model in the language of gauges and unit balls.

Proposition 3 (Gauge interpretation of the residual family)**.** *Let $q : \mathbb{R}^m \setminus \{0\} \to (0, \infty)$ be even and homogeneous of degree zero, and define $\rho(v) = \|v\|_2 q(v)$ for $v \neq 0$ and $\rho(0) = 0$. Then ρ is even and homogeneous of degree one. If, in addition, q is bounded above and below away from zero, the unit set*

$$B_\rho = \{v \in \mathbb{R}^m : \rho(v) \leq 1\} \tag{26}$$

is compact, absorbing, and centrally symmetric. Under these assumptions, ρ is a norm if and only if B_ρ is convex, equivalently if and only if ρ is subadditive.

This proposition explains the roles of the finite penalties. The homogeneity term encourages the residual factor to be stable along rays; the symmetrization enforces central symmetry; the subadditivity term is a sampled convexity test for the induced unit ball; and the reconstruction term checks whether the finite dyadic slope average is stable. The proposition also clarifies why the Euclidean anchor narrows the claim: the learned object is a bounded deformation of a known norm, not an arbitrary dissimilarity.

3.4. Finite Norm-Envelope Certificate

The gauge interpretation can be made constructive on a finite latent sample. Let $S \subset \mathbb{R}^m$ be a finite centrally symmetric set of nonzero latent differences and let $E_S = \operatorname{span}(S)$. For an even positive function $\rho : S \to (0, \infty)$, define the finite normalized atom set and its convex hull by

$$A_S(\rho) = \left\{ \pm \frac{v}{\rho(v)} : v \in S \right\}, \qquad B_S(\rho) = \operatorname{conv} A_S(\rho) \subset E_S. \tag{27}$$

The Minkowski functional of this finite unit ball is

$$p_S(x) = \inf\{t > 0 : x \in t B_S(\rho)\}, \qquad x \in E_S. \tag{28}$$

Equivalently,

$$p_S(x) = \inf\left\{ \sum_{j=1}^{r} |\alpha_j| \rho(u_j) : x = \sum_{j=1}^{r} \alpha_j u_j,\ u_j \in S,\ r \geq 1 \right\}. \tag{29}$$

Thus p_S is the finite sublinear envelope generated by the sampled values of ρ. It is computable by linear programming after fixing the atom set S.

Theorem 1 (Finite norm-envelope certificate)**.** *Let $S \subset \mathbb{R}^m$ be finite, centrally symmetric and spanning E_S, and let $\rho : S \to (0, \infty)$ be even. Then $B_S(\rho)$ is a compact, centrally symmetric convex body in E_S, and its Minkowski functional p_S is a norm on E_S. Moreover:*

(i) $p_S(u) \leq \rho(u)$ *for every $u \in S$.*

(ii) *p_S is the largest seminorm on E_S whose restriction to S is bounded above by ρ. In other words, if p is any seminorm with $p(u) \le \rho(u)$ for all $u \in S$, then $p(x) \le p_S(x)$ for all $x \in E_S$.*

(iii) *Define the finite-envelope gap*

$$\Gamma_S(\rho) = \sup_{u \in S} \left(1 - \frac{p_S(u)}{\rho(u)}\right)_+. \tag{30}$$

Then

$$(1 - \Gamma_S(\rho))\rho(u) \le p_S(u) \le \rho(u), \qquad u \in S. \tag{31}$$

(iv) *If, for some $\eta \in [0, 1)$, every decomposition $u = \sum_j \alpha_j u_j$ with $u, u_j \in S$ and real coefficients satisfies*

$$\rho(u) \le \frac{1}{1-\eta} \sum_j |\alpha_j| \rho(u_j), \tag{32}$$

then $\Gamma_S(\rho) \le \eta$.

This theorem turns the sampled subadditivity question into a finite certificate: if the finite-envelope gap is small, the learned dissimilarity admits an explicit norm p_S that approximates it on the sampled latent differences. The gap is one-sided because p_S is constructed as the largest norm lying below the observed values on S. A large gap indicates that the sampled unit ball contains nonconvex or non-norm-like radial features that are removed by convexification.

The finite dyadic estimator can now be related not only to the learned dissimilarity but also to the norm envelope.

Corollary 1 (Dyadic reconstruction relative to the finite envelope)**.** *Let the assumptions of Theorem 1 hold and let $v \in S$. Suppose that the normalized dyadic slopes $a_k(v) = 2^{-k}\rho(2^k v)$ are positive and satisfy Equation (23) for $k = 0, \dots, L-1$, where the values $\rho(2^k v)$ are defined on the ambient latent space. Then*

$$\frac{|\widehat{N}_L(v) - p_S(v)|}{\rho(v)} \le \frac{1}{L+1} \sum_{k=0}^{L} \left(e^{k\eta} - 1\right) + \Gamma_S(\rho). \tag{33}$$

Thus, dyadic homogeneity controls the reconstruction of ρ along sampled rays, while the envelope gap controls how far ρ is from an explicit finite norm on the sampled difference set. The bound is intentionally finite: it concerns the chosen sample S, the chosen dyadic range, and the ambient values of ρ at the scaled points.

The same finite construction also gives a clean asymptotic interpretation when one considers a sequence of increasingly rich finite certificates.

Theorem 2 (Asymptotic limit of finite certificates)**.** *Let E be a finite-dimensional real vector space. For each n, let $S_n \subset E$ be finite, centrally symmetric and spanning, and let p_n be the norm generated by Equation (28) from values $\rho_n : S_n \to (0, \infty)$. Assume that there are constants $0 < c \le C < \infty$ such that*

$$c\|x\|_2 \le p_n(x) \le C\|x\|_2, \qquad x \in E, \tag{34}$$

for all n. Then there is a subsequence (n_j) such that p_{n_j} converges locally uniformly on E to a norm p. Along any such convergent subsequence, if $\Gamma_{S_{n_j}}(\rho_{n_j}) \to 0$, then for every sequence $u_{n_j} \in S_{n_j}$ contained in a fixed compact subset and with $\inf_j \rho_{n_j}(u_{n_j}) > 0$,

$$\frac{|\rho_{n_j}(u_{n_j}) - p(u_{n_j})|}{\rho_{n_j}(u_{n_j})} \to 0. \tag{35}$$

This theorem is not used as an empirical claim about a single neural network. Its role is conceptual: it explains how a sequence of finite certificates with vanishing envelope gaps can converge to an actual norm, thereby connecting the finite procedure back to the AMN intuition without asserting global normability from one finite experiment.

3.5. Refinement Towers and Finite AMN Limits

The envelope construction also allows a closer finite analog of AMN norm reconstruction. In the asymptotic AMN picture, a norm emerges as a limit of rescaled metric information. In the neural finite setting, the corresponding object is not a single limit in scale only, but a sequence of refined finite certificates: more latent differences, more probed scales, deeper dyadic averages, or improved residual heads produce a succession of witness norms.

Definition 1 (Finite AMN-refinement tower)**.** *Fix a finite-dimensional real vector space E. A finite AMN-refinement tower is a sequence*

$$\mathcal{C}_n = (S_n, \rho_n, L_n, \Lambda_n, p_n, \Gamma_n), \tag{36}$$

where $S_n \subset E$ is finite, centrally symmetric and spanning, $\rho_n : S_n \to (0, \infty)$ is even, $p_n = p_{S_n,\rho_n}$ is the norm-envelope witness defined by Equation (28), and $\Gamma_n = \Gamma_{S_n}(\rho_n)$ is the finite-envelope gap. The tower is uniformly anchored if there exists $0 < c \le C < \infty$ such that

$$c\|x\|_2 \le p_n(x) \le C\|x\|_2, \qquad x \in E, \tag{37}$$

for all n. For a compact set $K \subset E$, define the refinement gap

$$\Xi_n(K) = \sup_{x \in K} |p_{n+1}(x) - p_n(x)|. \tag{38}$$

The following theorem records the monotone situation in which the dissimilarity is fixed, and only the sampled latent directions are refined. This is the cleanest finite analog of adding more observed directions to an approximate unit ball.

Theorem 3 (Monotone refinement for a fixed dissimilarity)**.** *Let $S_1 \subset S_2 \subset \cdots \subset E$ be finite, centrally symmetric, and spanning, and let ρ be an even positive function on $S_\infty = \bigcup_n S_n$. Define*

$$B_n = \operatorname{conv}\left\{ \pm \frac{v}{\rho(v)} : v \in S_n \right\}, \qquad p_n = \gamma_{B_n}, \tag{39}$$

and

$$B_\infty = \overline{\operatorname{conv}}\left\{ \pm \frac{v}{\rho(v)} : v \in S_\infty \right\}. \tag{40}$$

Assume that B_∞ is a compact convex body in E. Then $B_n \subseteq B_{n+1} \subseteq B_\infty$, the gauges satisfy

$$p_{n+1}(x) \le p_n(x), \qquad x \in E, \tag{41}$$

B_n converges to B_∞ in Hausdorff distance, and p_n converges locally uniformly to the norm $p_\infty = \gamma_{B_\infty}$. If, in addition, $\Gamma_{S_n}(\rho|_{S_n}) \to 0$, then for every sequence $u_n \in S_n$ contained in a fixed compact subset and satisfying $\inf_n \rho(u_n) > 0$,

$$\frac{|\rho(u_n) - p_\infty(u_n)|}{\rho(u_n)} \to 0. \tag{42}$$

For trained neural models, the dissimilarity may also change between refinement levels. In that case, exact monotonicity is replaced by a Cauchy condition on the witness norms.

Theorem 4 (Coherent finite AMN-refinement limit)**.** *Let $\mathcal{C}_n = (S_n, \rho_n, L_n, \Lambda_n, p_n, \Gamma_n)$ be a uniformly anchored finite AMN-refinement tower in a fixed finite-dimensional space E. Suppose that for every compact set $K \subset E$,*

$$\sum_{n=1}^{\infty} \Xi_n(K) < \infty. \tag{43}$$

Then p_n converges locally uniformly to a norm p_∞ on E. For $v \in S_n$ with $\rho_n(v) > 0$, define the relative dyadic reconstruction residual

$$\Theta_n(v) = \frac{|\widehat{N}^{(n)}_{L_n}(v) - \rho_n(v)|}{\rho_n(v)}, \tag{44}$$

where $\widehat{N}^{(n)}_{L_n}$ is computed from ρ_n. If $u_n \in S_n$ remains in a fixed compact set, $\inf_n \rho_n(u_n) > 0$, and

$$\Gamma_n + \Theta_n(u_n) \to 0, \tag{45}$$

then

$$\frac{|\rho_n(u_n) - p_\infty(u_n)|}{\rho_n(u_n)} \to 0, \qquad \frac{|\widehat{N}^{(n)}_{L_n}(u_n) - p_\infty(u_n)|}{\rho_n(u_n)} \to 0. \tag{46}$$

The tower formulation identifies the finite analog of the AMN limiting picture used in this paper. The asymptotic AMN construction reconstructs a norm from the large-scale behavior of one metric; here, a norm-like latent geometry is obtained as the local uniform limit of finite witness norms, provided the sampled envelopes are coherent across refinements. The theorem is conditional: it does not assert that training automatically produces such a tower. It states what additional coherence would be needed to interpret successive finite audits as convergence toward a limiting norm.

Remark 2 (Collapsed directions and quotient geometry)**.** *If the lower bound in Equation (37) is dropped, a locally uniform limit of the witness norms may be a seminorm. Its null space*

$$E_0 = \{x \in E : p_\infty(x) = 0\} \tag{47}$$

then defines a quotient norm on E/E_0. This is a natural finite counterpart of the AMN intuition that bounded or collapsed directions can be factored out before a norm-like geometry is recovered. In representation-learning terms, E_0 corresponds to latent directions that the learned dissimilarity does not separate at the limiting resolution of the tower.

3.6. Stable Kernel-Derived Distance Quotient Model

The quotient interpretation in Section 2.2 can be stated as a simple model result. It is not a neural training theorem; rather, it records the AMN large-scale mechanism

behind exponential similarities whose logarithmic, or square-root logarithmic, transform is a bounded perturbation of a linear seminorm.

Proposition 4 (Stable kernel-derived distance and quotient latent geometry)**.** *Let E be a finite-dimensional real vector space, let H be a normed vector space, let $A : E \to H$ be linear, and let $D : E \times E \to \mathbb{R}_+$ be a symmetric distance-like quantity satisfying, for some $B < \infty$,*

$$|D(x,y) - \|A(x-y)\|_H| \le B, \qquad x, y \in E. \tag{48}$$

Define the large-scale slope and bounded-direction set by

$$\nu(v) = \lim_{n\to\infty} \frac{D(nv,0)}{n}, \qquad E_0 = \{v \in E : \sup_{n\ge 1} D(nv,0) < \infty\}. \tag{49}$$

Then the limit exists for every $v \in E$ and

$$\nu(v) = \|Av\|_H, \qquad E_0 = \ker A. \tag{50}$$

Consequently, ν induces a norm on the quotient E/E_0 by $\nu([v]) = \|Av\|_H$.

The proof is immediate and is included in Appendix A. If a learned exponential similarity has Laplacian form $K(x,y) = \exp(-D(x,y)/\tau)$ and the logarithmic quantity $D = -\tau \log K$ satisfies Equation (48), then AMN scaling recovers the stable quotient distance ν. For Gaussian similarities, the same statement applies to the square-root log-distance in Equation (11), while $-2\ell^2 \log K$ itself should be read as a squared distance-like quantity. The directions in E_0 are precisely those that the kernel-derived distance cannot distinguish at a large scale. The statement does not assert that an arbitrary learned kernel yields a metric: symmetry, separation, and triangle-type properties must still be checked or regularized. This proposition is the clean model case that the finite norm-envelope and refinement-tower diagnostics approximate on sampled latent differences.

Corollary 2 (Specific-energy stability and affinity consistency)**.** *Let $D, D' : E \times E \to \mathbb{R}_+$ be two distance-like quantities satisfying $|D'(x,y) - D(x,y)| \le B$ for all x, y and some $B < \infty$. If*

$$\nu_D(v) = \lim_{n\to\infty} \frac{D(nv,0)}{n} \tag{51}$$

exists, then $\nu_{D'}(v)$ exists and equals $\nu_D(v)$. In addition, if D is subadditive on triples,

$$D(x,z) \le D(x,y) + D(y,z), \tag{52}$$

then the exponential affinity $K_D(x,y) = \exp(-D(x,y)/\tau)$ satisfies

$$K_D(x,z) \ge K_D(x,y) K_D(y,z). \tag{53}$$

This corollary gives the precise sense in which the energy language is useful here. Bounded perturbations conserve the AMN-specific energy, while subadditivity of the distance-like energy becomes multiplicative consistency of the corresponding bounded affinity. The statement is not a physical conservation law and does not make $-\log K$ an entropy. Entropic quantities enter only after normalization of affinities into probabilities, as described in Section 2.3.

Finally, the dyadic grid used in training does not by itself control all intermediate scales. The following conditional statement records the additional regularity needed to transfer dyadic control to off-grid scales.

Lemma 1 (Intermediate-scale control under log-scale regularity)**.** *Fix $v \neq 0$ and define $h_v(t) = e^{-t}\rho(e^t v)$ on an interval $I \subset \mathbb{R}$. Let $G = \{t_0, \dots, t_J\} \subset I$ be a grid such that every $t \in I$ is within distance at most $\Delta/2$ of some grid point. If h_v is M-Lipschitz on I and*

$$\max_{t_j \in G} |h_v(t_j) - h_v(0)| \leq \alpha, \tag{54}$$

then every $t \in I$ satisfies

$$|h_v(t) - h_v(0)| \leq \alpha + M\Delta/2. \tag{55}$$

The lemma is not used as a training guarantee because the Lipschitz constant of the trained neural head is not certified in the experiments. Its purpose is to identify the missing regularity assumption behind off-dyadic generalization and to motivate the supplementary unseen-scale audits.

Corollary 3 (From expected losses to sampled finite defects)**.** *Let*

$$H(v, \lambda) = \left| \frac{d_\theta(\lambda v)}{|\lambda| d_\theta(v) + \varepsilon} - 1 \right|, \tag{56}$$

$$S(v, w) = \frac{[d_\theta(v + w) - d_\theta(v) - d_\theta(w)]_+}{d_\theta(v) + d_\theta(w) + \varepsilon}, \tag{57}$$

$$R(v) = \frac{|d_\theta(v) - \widehat{N}_L(v)|}{d_\theta(v) + \varepsilon}. \tag{58}$$

If the squared training losses satisfy $\mathbb{E}H^2 \leq \ell_h$, $\mathbb{E}S^2 \leq \ell_s$ and $\mathbb{E}R^2 \leq \ell_r$ under the sampling distributions used in training, then

$$\mathbb{E}H \leq \sqrt{\ell_h}, \qquad \mathbb{E}S \leq \sqrt{\ell_s}, \qquad \mathbb{E}R \leq \sqrt{\ell_r}, \tag{59}$$

and, for any $t > 0$,

$$\mathbb{P}(H > t) \leq \frac{\ell_h}{t^2}, \qquad \mathbb{P}(S > t) \leq \frac{\ell_s}{t^2}, \qquad \mathbb{P}(R > t) \leq \frac{\ell_r}{t^2}. \tag{60}$$

In particular, with probability at least $1 - \ell_s/t^2$ over a sampled pair (v, w),

$$d_\theta(v + w) \leq d_\theta(v) + d_\theta(w) + t\big(d_\theta(v) + d_\theta(w) + \varepsilon\big). \tag{61}$$

The corollary is intentionally elementary: it uses Jensen's inequality and Markov's inequality. Its purpose is to close the gap between the pointwise assumptions of Proposition 1 and the minibatch expected losses optimized by the algorithm. It does not turn the empirical regularizer into a global metric certificate.

Corollary 4 (Empirical concentration for reported finite defects)**.** *Fix a trained model and one diagnostic distribution, for instance, the training-gallery or test-query latent-difference distribution. Let $X_1, \dots, X_M$ be independent draws of one of the non-negative normalized defects H, S or R in*

Equations (56)–(58). *If* $0 \le X_i \le B$ *almost surely and* $\widehat{\mu}_M = M^{-1}\sum_{i=1}^{M} X_i$, *then Hoeffding's inequality [18] gives, for every* $\eta > 0$,

$$\mathbb{P}(|\widehat{\mu}_M - \mathbb{E}X| > \eta) \le 2\exp\left(-\frac{2M\eta^2}{B^2}\right). \tag{62}$$

Consequently, the diagnostic estimates reported from $M = 4096$ *sampled differences should be interpreted as finite-sample estimates of the corresponding sampled-distribution defects, conditional on the trained model and the chosen latent distribution.*

This concentration statement is also deliberately modest and follows from Hoeffding's inequality [18]. It addresses Monte Carlo error in reporting finite defects after a model has been trained, while learning-theoretic generalization of the neural encoder and held-out validation remain separate issues.

3.7. Mathematical Contribution and Limitations

The primary contribution is a finite framework for diagnosing and regularizing norm-like geometry in learned latent dissimilarities. It has three parts. First, the diagnostics D_{hom}, D_{sub}, and D_{rec} measure finite violations of scaling, subadditivity, and dyadic reconstruction on sampled latent differences. Second, the theoretical statements explain what those diagnostics control in the finite setting: stability of the dyadic estimator, a gauge and unit-ball interpretation, finite norm-envelope witnesses, refinement limits under explicit coherence assumptions, and quotient behavior for stable kernel-derived distances. Third, the experiments test whether adding the corresponding penalties improves finite geometry inside a Euclidean-anchored residual family.

The limitations are equally important. The results do not prove AMN rigidity, global normability, convergence of $\widehat{N}_L$ as $L \to \infty$, or a metric certificate on the original input space X. They also do not show that the residual family is a better retrieval model than Euclidean supervised-contrastive or proxy-based objectives. The full-split experiments deliberately include those Euclidean baselines to show that the proposed framework is about finite geometric control and auditability, not retrieval leadership.

The hyperparameter protocol should be read in this light. The AMN weights were selected in exploratory runs and then kept fixed for the final controlled audits. No additional validation split was introduced beyond the held-out evaluation subsets and official test sets. Therefore, the reported numerical results are best interpreted as a reproducible finite-geometry audit with fixed choices rather than as an optimized model-selection claim. The post hoc shrinkage, residual-scale, and off-grid audits were added to test alternative explanations and were not used to tune the main reported models.

3.8. Synthetic Diagnostic Example

Before using neural networks, it is useful to inspect the diagnostics on explicit functions. We evaluate six functions in $\mathbb{R}^8$: the Euclidean norm, an anisotropic norm $v \mapsto \sqrt{v^\top A v}$ with positive diagonal A, a superlinear radial perturbation $\rho(v) = \|v\|_2 + 0.08\|v\|_2^2$, a bounded log-periodic radial perturbation, a homogeneous but nonconvex gauge $\rho(r,\varphi) = r(1 + 0.35\cos 4\varphi)$ on the first two coordinates, and a dyadic-aliasing perturbation $\rho(v) = \|v\|_2\{1 + 0.18\sin(2\pi\log_2\|v\|_2)\}$ away from the origin. For any exact norm, $2^{-k}\rho(2^k v) = \rho(v)$ for all k, so $\widehat{N}_L(v) = \rho(v)$ and the three finite diagnostics vanish up to numerical precision. The homogeneous nonconvex gauge shows why homogeneity and reconstruction are not enough without subadditivity. The dyadic-aliasing perturbation shows why off-grid scale audits are informative: it can look stable on dyadic scales while varying at intermediate scales.

Tables 4–6 and Figure 2 illustrate the finite mechanism behind Propositions 1–3 and Theorems 1 and 3. Exact norms are reconstructed exactly by the finite dyadic estimator. The quadratic perturbation has rapidly changing normalized slopes and a large reconstruction defect. The homogeneous nonconvex gauge has negligible homogeneity and reconstruction defects but a nonzero subadditivity defect and a positive finite-envelope gap, showing why sampled convexity is necessary. The dyadic-aliasing perturbation has nearly zero dyadic homogeneity, reconstruction, and envelope defects on sampled dyadic directions but a visible off-grid homogeneity defect, motivating the unseen-scale diagnostics described in the supplementary scripts. The refinement demonstration shows the other direction of the theory: for an exact norm, finite sampled atom hulls give polygonal witness norms that converge as directions are refined. Thus, the synthetic examples support the interpretation of D_{hom}, D_{sub}, and D_{rec} as complementary finite diagnostics of norm-like behavior rather than as arbitrary regularization terms.

Table 4. Synthetic finite diagnostics in $\mathbb{R}^8$. Values are Monte Carlo averages over 20,000 Gaussian pairs, with $L = 4$. The dyadic homogeneity column uses the training scales $\{1/2, 2, 4\}$; the off-grid column uses $\{\sqrt{2}, 1.5, 3, 6\}$.

Function	Role	D_{hom}	D_{sub}	D_{rec}	Off-Grid D_{hom}	Slope Span
Euclidean norm	exact norm	0.0000	0.0000	0.0000	0.0000	0.0000
Anisotropic norm	exact non-Euclidean norm	0.0000	0.0000	0.0000	0.0000	0.0000
Quadratic radial perturbation	superlinear non-norm	0.2673	0.0029	0.9265	0.3525	2.6726
Shifted log-periodic perturbation	bounded scale oscillation	0.1889	0.0088	0.1536	0.1960	0.2680
Homogeneous nonconvex gauge	homogeneous nonconvex gauge	0.0000	0.0240	0.0000	0.0000	0.0000
Dyadic-aliasing perturbation	dyadic trap	0.0000	0.0060	0.0000	0.2297	0.0000

Table 5. Finite norm-envelope diagnostic on 96 directions in $\mathbb{R}^2$. The envelope gap is zero for sampled directions of exact norms, positive for a homogeneous nonconvex gauge, and does not detect the dyadic-aliasing radial perturbation unless off-grid scales are evaluated.

Function	ρ Range on S	Mean Γ Contribution	Max Γ	Dyadic Hom.	Off-Grid Hom.
Euclidean norm	[1.000, 1.000]	0.0000	0.0000	0.0000	0.0000
Anisotropic norm	[1.000, 1.581]	0.0000	0.0000	0.0000	0.0000
Homogeneous nonconvex gauge	[0.620, 1.380]	0.2020	0.3900	0.0000	0.0000
Dyadic-aliasing radial perturbation	[1.000, 1.000]	0.0000	0.0000	0.0000	0.0458

Table 6. Finite norm-refinement tower demonstration in $\mathbb{R}^2$. Increasing the number of sampled directions refines the convex atom hull and decreases the witness-gauge overestimate on a dense probe grid. The refinement gap is the maximum change relative to the previous level.

Target Norm	Directions	Atoms	Mean Rel. Error	Max Rel. Error	Refinement Gap
Euclidean	8	16	0.0130	0.0196	–
Euclidean	16	32	0.0032	0.0048	0.0096
Euclidean	32	64	0.0008	0.0012	0.0024
Euclidean	64	128	0.0002	0.0003	0.0006
anisotropic	8	16	0.0162	0.0629	–
anisotropic	16	32	0.0040	0.0181	0.0314
anisotropic	32	64	0.0010	0.0047	0.0085
anisotropic	64	128	0.0003	0.0011	0.0022

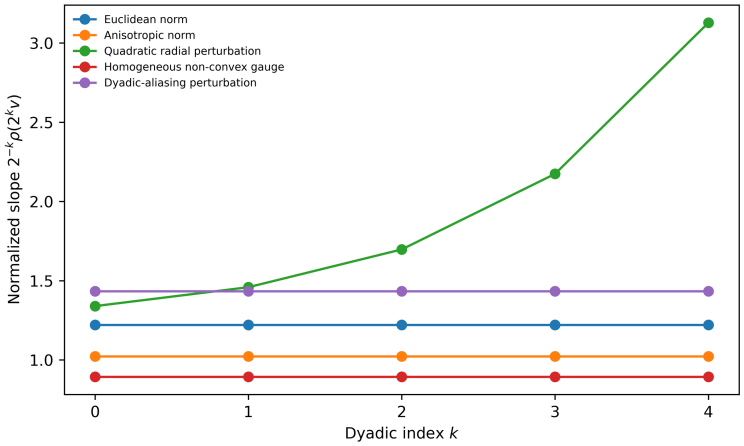


Figure 2. Synthetic normalized dyadic slopes for representative explicit functions. Exact norms have constant normalized slopes. Scale-dependent perturbations produce finite-scale slope variation; the dyadic-aliasing perturbation illustrates that dyadic scales alone may miss intermediate-scale oscillations.

3.9. Training Objective and Variants

Definition 2 (Training objective and finite AMN-inspired regularizer)**.** *Let $\mathcal{T}$ be the empirical distribution of anchor-positive-negative triples (x_a, x_p, x_n), where x_a is an anchor observation, x_p has the same class label as x_a, and x_n has a different class label. Let $z_a = f_\psi(x_a)$, $z_p = f_\psi(x_p)$ and $z_n = f_\psi(x_n)$. For margin $M > 0$, define the task loss by*

$$\mathcal{L}_{\text{task}} = \mathbb{E}_{\mathcal{T}}\left[d_\theta(z_a - z_p)^2 + [M - d_\theta(z_a - z_n)]_+^2\right]. \tag{63}$$

Let $v, w \sim \mathcal{V}_B$ be independent minibatch latent differences and let $\lambda \sim \text{Unif}(\Lambda)$ be independent of them. With this convention, the squared AMN-inspired loss terms are

$$\mathcal{L}_{\text{hom}} = \mathbb{E}_{\substack{v\sim\mathcal{V}_B \\ \lambda\sim\text{Unif}(\Lambda)}}\left(\frac{d_\theta(\lambda v)}{|\lambda| d_\theta(v) + \varepsilon} - 1\right)^2, \tag{64}$$

$$\mathcal{L}_{\text{sub}} = \mathbb{E}_{\substack{v,w\sim\mathcal{V}_B \\ \text{independent}}}\left(\frac{[d_\theta(v+w) - d_\theta(v) - d_\theta(w)]_+}{d_\theta(v) + d_\theta(w) + \varepsilon}\right)^2, \tag{65}$$

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{v\sim\mathcal{V}_B}\left(\frac{d_\theta(v) - \widehat{N}_L(v)}{d_\theta(v) + \varepsilon}\right)^2. \tag{66}$$

The optional symmetry and non-degeneracy terms are

$$\mathcal{L}_{\text{sym}} = \mathbb{E}_{v\sim\mathcal{V}_B}\frac{|\rho_\theta(v) - \rho_\theta(-v)|}{\rho_\theta(v) + \rho_\theta(-v) + \varepsilon}, \tag{67}$$

$$\mathcal{L}_{\text{nd}} = \mathbb{E}_{v\sim\mathcal{V}_B}\left[\left(\frac{[\alpha_{\min}\|v\|_2 - d_\theta(v)]_+}{\alpha_{\min}\|v\|_2 + \varepsilon}\right)^2 + \left(\frac{[d_\theta(v) - \alpha_{\max}(\|v\|_2 + \varepsilon)]_+}{\alpha_{\max}\|v\|_2 + \varepsilon}\right)^2\right], \tag{68}$$

where $\alpha_{\min}$ and $\alpha_{\max}$ are lower and upper slope bounds. For non-negative weights w_{hom}, w_{sub}, w_{rec}, w_{sym} and w_{nd}, define the total training objective by

$$\mathcal{L} = \mathcal{L}_{\text{task}} + w_{\text{hom}}\mathcal{L}_{\text{hom}} + w_{\text{sub}}\mathcal{L}_{\text{sub}} + w_{\text{rec}}\mathcal{L}_{\text{rec}} + w_{\text{sym}}\mathcal{L}_{\text{sym}} + w_{\text{nd}}\mathcal{L}_{\text{nd}}. \tag{69}$$

In all final reported experiments, $M = 1$, $w_{\text{sym}} = w_{\text{nd}} = 0$, $\alpha_{\min} = 0.02$, $\alpha_{\max} = 50$, $\Lambda = \{1/2, 2, 4\}$ and $L = 4$. The two Fashion-MNIST AMN-inspired variants differ in the three main weights:

- **AMN-inspired balanced**: $w_{\text{hom}} = 0.06$, $w_{\text{sub}} = 0.12$, $w_{\text{rec}} = 0.005$;
- **AMN-inspired geom-first**: $w_{\text{hom}} = 0.09$, $w_{\text{sub}} = 0.08$, $w_{\text{rec}} = 0.01$.

We compare them against a Euclidean triplet baseline, a residual task-only dissimilarity with $w_{\text{hom}} = w_{\text{sub}} = w_{\text{rec}} = 0$, and non-AMN controls: stronger weight decay, a generic smoothness penalty, reconstruction-only regularization, and homogeneity plus subadditivity without the explicit reconstruction term. The UCI HAR corroboration uses the same objective, but only the Euclidean, residual task-only, and AMN-inspired balanced families are compared; because it is a small CPU secondary run rather than the Fashion-MNIST controlled experiment, the AMN weights are set to $w_{\text{hom}} = 0.12$, $w_{\text{sub}} = 0.24$, and $w_{\text{rec}} = 0.02$.

The Euclidean triplet baseline is included as a standard norm-constrained reference, and the residual task-only model is the main within-family comparator. We also executed a full official Fashion-MNIST split audit with supervised-contrastive and proxy-anchor-style Euclidean baselines. These objectives change the training criterion and are therefore not direct tests of the AMN-inspired penalties, but they give an explicit reference for retrieval performance under modern norm-like embeddings.

Hyperparameter Selection

The AMN weights were selected in an exploratory Fashion-MNIST grid and then fixed before the final ten-seed controlled run, the non-AMN control run, and the full official split audit. The exploratory range was $w_{\text{hom}} \in \{0.03, 0.06, 0.09\}$, $w_{\text{sub}} \in \{0.04, 0.08, 0.12\}$ and $w_{\text{rec}} \in \{0, 0.005, 0.01\}$ on a reduced candidate sweep. Selection used a scalar score that favored defect reductions while penalizing Recall@1 loss relative to the residual task-only run. No separate validation split was introduced beyond the held-out evaluation subset; therefore, these weights are reported as pre-fixed exploratory choices rather than as statistically optimized hyperparameters, and the full official split should be read as a fixed-weight audit rather than as a validation-selected benchmark. Appendix B provides the compact hyperparameter and experimental-protocol details underlying these settings; Table A2 reports a representative sensitivity excerpt from the archived grid.

3.10. Data Sets, Implementation and Evaluation Protocol

3.10.1. Primary Benchmark: Fashion-MNIST

The main experiments use Fashion-MNIST [19], a public benchmark of Zalando article images. The dataset contains 70,000 grayscale images of size 28×28: 60,000 training images and 10,000 test images. Each image belongs to one of ten classes: T-shirt/top, trousers, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. Figure 3 shows representative images from the public Fashion-MNIST visualization. The benchmark was chosen because it keeps the simple size and split structure of MNIST [20] but replaces handwritten digits with visually more heterogeneous product images, making it a controlled yet nontrivial test bed for retrieval and metric-learning behavior.

For computational control and paired multi-seed auditing, the controlled Fashion-MNIST experiments use ten random seeds, stratified training subsets of 20,000 examples and test subsets of 5000 examples. Retrieval evaluation uses up to 3500 test queries against a 16,000-example training gallery. These limits were chosen to make the residual variants, non-AMN controls, robustness probes, and paired statistics comparable under a fixed compute budget. To address full-split reproducibility and modern-baseline context, we additionally ran the complete official 60,000/10,000 split over ten seeds for Euclidean triplet, supervised contrastive, proxy-anchor-style, residual task-only, AMN-balanced, and AMN-geometry-first variants. In this full-split audit, all 10,000 test examples are queries and the 60,000 training examples form the gallery; reconstructed-distance retrieval is skipped

for tractability, but the finite reconstruction defect D_{rec} is still computed. The encoder is a compact convolutional neural network producing a latent embedding of dimension 32. Training and evaluation were implemented in PyTorch 2.7.1+cu118 [21].

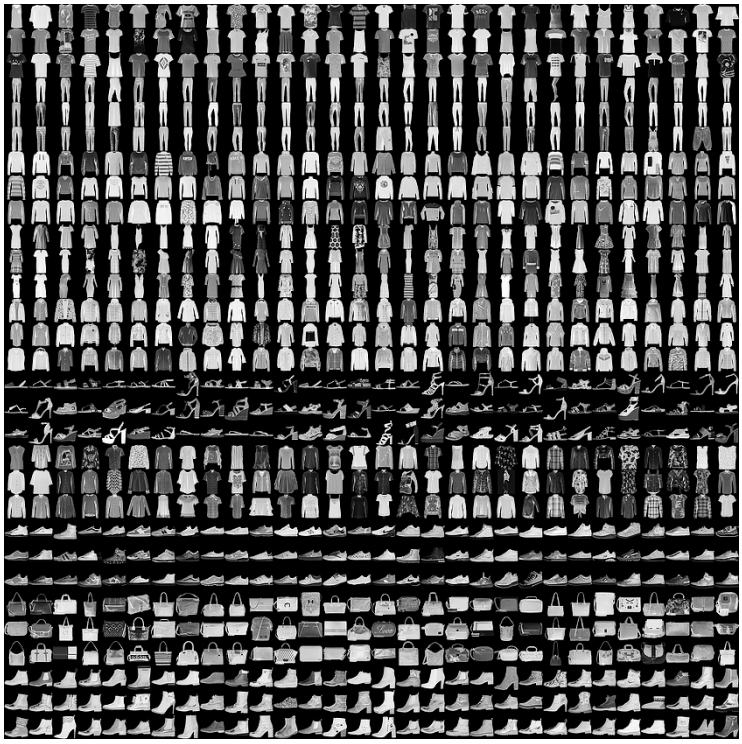


Figure 3. Representative Fashion-MNIST images from the public project visualization. The classes include clothing items and accessories such as tops, trousers, shoes, bags, and ankle boots. Source: Fashion-MNIST project repository, MIT license.

3.10.2. Secondary Non-Visual Benchmark: UCI HAR

The main cross-domain corroboration uses the UCI Human Activity Recognition Using Smartphones dataset (UCI HAR) [22,23]. The data consist of smartphone accelerometer and gyroscope measurements collected from 30 volunteers performing six activities: walking, walking upstairs, walking downstairs, sitting, standing, and laying. The public benchmark provides an official subject-based split of 7352 training records and 2947 test records, and represents each window by a 561-dimensional feature vector derived from 2.56 s inertial-sensor windows sampled at 50 Hz with 50% overlap. This benchmark is non-visual: the inputs are not pixels but standardized sensor-derived feature vectors. It was chosen because it keeps the supervised retrieval setting and class-label structure while testing the AMN-inspired mechanism on tabular time-series-derived data rather than images.

In the UCI HAR run, the official 561-feature vectors are standardized using the training-set mean and standard deviation. The encoder is a compact multilayer perceptron with hidden dimension 64 and latent dimension 32. The official train/test split is fixed; the reported seeds affect initialization and minibatch sampling, not the subject split. The non-visual corroboration uses ten seeds and the full official evaluation geometry: all 2947 test records are queries and all 7352 training records form the gallery. Reconstructed-distance retrieval is skipped, and only the diagnostic reconstruction defect D_{rec} is reported. The UCI HAR experiment remains secondary because it uses a compact CPU protocol and no separate hyperparameter search. Figure 4 gives a two-dimensional PCA visualization of the standardized sensor features.

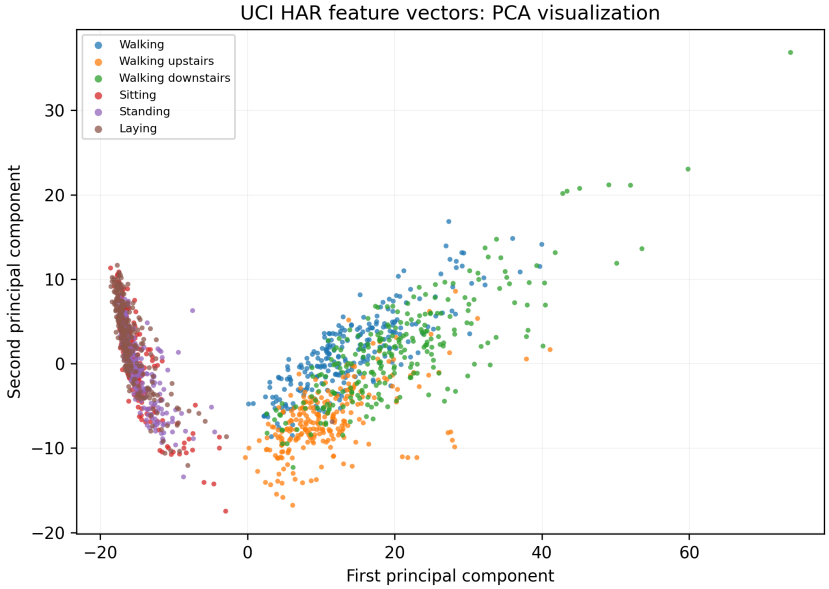


Figure 4. PCA visualization of representative UCI HAR records after standardizing the 561-dimensional sensor-feature vectors. The plot is not used for training; it illustrates that the non-visual benchmark has class structure in tabular inertial-sensor features. Source: UCI HAR public dataset.

3.10.3. Auxiliary Image-Domain Check: Scikit-Learn Digits

We retain a smaller auxiliary experiment on the handwritten Digits dataset distributed with scikit-learn. This dataset contains 1797 images of handwritten digits, each represented as an 8×8 grayscale image with integer pixel values in the range 0–16, and has ten classes corresponding to the digits 0–9 [24]. Representative examples are shown in Figure 5. Digits is no longer the main secondary benchmark, because it is still an image dataset; it is kept only as a small image-domain mechanism check.

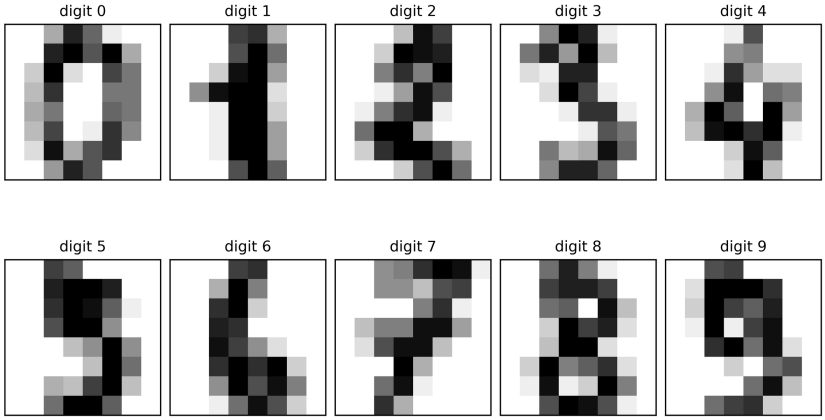


Figure 5. Representative examples from the scikit-learn Digits dataset. The images are 8×8 handwritten-digit thumbnails; the displayed examples use the first available sample from each class after loading the dataset.

For the Digits pilot, the 8×8 images are rescaled to $[0, 1]$, upsampled by nearest-neighbor replication to 24×24, and padded to 28×28 so that the same convolutional code path can be reused. The supplementary scripts generate deterministic stratified train/test splits with a default test fraction of 0.25. The archived pilot summaries use the same residual family and finite diagnostics as the Fashion-MNIST study, but they are interpreted only as an auxiliary check of the mechanism.

3.10.4. Evaluation and Statistical Protocol

Retrieval metrics are computed using each test example as a query and ranking training-gallery examples by increasing learned dissimilarity. Hence, the query itself is not present in the gallery. For $k \geq 1$, Recall@k is defined as

$$\text{Recall@}k = \frac{1}{|Q|} \sum_{q \in Q} \mathbf{1}\{\text{at least one of the top } k \text{ retrieved items has label } y_q\}, \quad (70)$$

where Q is the query set and y_q is the class label of query q. In particular, Recall@1 is nearest-neighbor retrieval accuracy under this label criterion. We also report nearest-neighbor accuracy and mean average precision where applicable. Geometry is evaluated by D_{hom}, D_{sub}, D_{rec}, the dyadic homogeneity defects at 2 and 4, and the magnitude and rate of triangle violations.

Unless explicitly stated otherwise, reported values are means and standard deviations over seeds. Fashion-MNIST paired comparisons are computed by pairing variants with the residual task-only baseline at the same seed. We report descriptive paired differences, paired t-tests, and Wilcoxon signed-rank tests in the Supplementary Material (CSV files); the main text uses these values only to indicate seed-level consistency of the defect reductions. The full official split audit uses ten seeds for all six variants. The additional Fashion-MNIST theory audits use three seeds and are reported as targeted diagnostics rather than as new model-selection sweeps. UCI HAR uses ten seeds on the fixed official subject split and is interpreted as secondary corroboration, not as a broad non-visual benchmark claim. Robustness probes for the image benchmark are included for Gaussian noise, central occlusion, contrast perturbation, and pixel shift; they are secondary diagnostics and are not used to claim universal robustness. Table 7 summarizes the main training, evaluation, and reproducibility protocol.

Table 7. Training, evaluation and reproducibility protocol. The table lists the details most relevant for auditing the reported experiments.

Item	Protocol Detail
Optimizer	AdamW in both pretraining and residual fine-tuning; no learning-rate schedule and no early stopping.
Fashion-MNIST controlled training	10 seeds; stratified 20,000-example training subsets and 5000-example test subsets; 14 pretraining rounds and 8 fine-tuning rounds; each round uses 70 fixed minibatch steps rather than a full pass over all 60,000 examples; batch size 384.
Fashion-MNIST full-split audit	10 seeds on the official 60,000/10,000 split; Euclidean triplet, supervised-contrastive, proxy-anchor-style, residual task-only, AMN-balanced and AMN-geometry-first variants; 14 pretraining rounds and 8 fine-tuning rounds for triplet/residual variants; each round uses 70 fixed minibatch steps; batch size 384; reconstructed retrieval skipped, diagnostic D_{rec} retained.
Fashion-MNIST controlled evaluation	3500 test queries ranked against a 16,000-example training gallery; the query image is never part of the gallery. Retrieval is by increasing learned dissimilarity.
Fashion-MNIST full-split evaluation	all 10,000 test queries ranked against the 60,000-example training gallery; the query image is never part of the gallery.
Fashion-MNIST diagnostics	4096 sampled training-gallery latent differences per seed in both the controlled and full-split protocols; dyadic depth $L = 4$ and scales $\{1/2, 2, 4\}$.
UCI HAR training/evaluation	Official subject train/test split; ten seeds affect initialization and minibatch sampling; 4 pretraining rounds and 8 fine-tuning rounds; each round uses 30 fixed minibatch steps; batch size 128; all 2947 test records are queries, and all 7352 training records form the gallery. The subject split is fixed, and the reported seeds change only in initialization and minibatch order.

Table 7. *Cont.*

Item	Protocol Detail
Triplet sampling	Anchors are sampled uniformly from the training split; positives are sampled from the same class, avoiding the anchor when possible; negatives are sampled from a uniformly chosen different class.
Image encoder	Compact CNN: two 3×3 convolution–batch-normalization–ReLU blocks at width 32, max pooling, two blocks at width 64, max pooling, one block at width 128, adaptive average pooling, dropout 0.05, and a linear 32-dimensional embedding.
UCI HAR encoder	Two-layer MLP on standardized 561-dimensional feature vectors, hidden width 64, layer normalization, ReLU, dropout 0.05, and a 32-dimensional embedding.
Residual head g_θ	MLP applied to $[u, u^2, \log(1+\|v\|_2), \|v\|_2/(1+\|v\|_2)]$, where $u = v/(\|v\|_2 + 10^{-7})$; hidden width 128 for Fashion-MNIST and 64 for UCI HAR; final layer initialized at zero.
Hyperparameter selection	AMN weights were selected in exploratory runs before the final seed sweep and then held fixed for the controlled and full-split evaluations; no seed-specific tuning and no full-split test-based tuning were used. The explored grid is summarized in Table A2.
Compute cost	Full-split Fashion-MNIST means on one Tesla V100-SXM2 32 GB GPU: about 18–20 s/seed for supervised contrastive and proxy-anchor-style baselines, 55 s/seed for Euclidean triplet, and 95–96 s/seed for residual/AMN variants under the reported full-split protocol. The additional Fashion-MNIST audits use three seeds on the same V100. UCI HAR full-query/full-gallery CPU means are about 7 s/seed for Euclidean and 61 s/seed for residual/AMN variants.

4. Results

The empirical analysis is organized around the geometric claim rather than around a state-of-the-art retrieval claim. The key comparison is within the same residual family: task-only versus AMN-inspired variants. Euclidean triplet learning is reported as a standard norm-like reference, and the control regularizers test whether the observed changes are due merely to additional regularization.

4.1. Fashion-MNIST Controlled Experiment

Table 8 reports the controlled Fashion-MNIST results. All entries are means and standard deviations over ten paired seeds. The Euclidean triplet baseline is nearly defect-free because it is a norm in the learned embedding. The residual task-only model gives the highest Recall@1 among the residual models, but it has large finite homogeneity and reconstruction defects. AMN-inspired balanced changes Recall@1 from 0.8769 ± 0.0070 to 0.8749 ± 0.0056 while reducing D_{hom} from 0.2634 ± 0.0098 to 0.1587 ± 0.0056, D_{sub} from 0.0152 ± 0.0018 to 0.0084 ± 0.0013, and D_{rec} from 0.2801 ± 0.0103 to 0.1524 ± 0.0064. AMN-inspired geom-first further reduces the defects at a slightly larger retrieval cost.

Table 8. Fashion-MNIST controlled experiment. Values are the mean ± standard deviation over ten seeds. The Euclidean triplet baseline is a standard triplet-trained embedding with Euclidean latent distance.

Variant	Recall@1	Recall@1($\widehat{N}_L$)	D_{hom}	D_{sub}	D_{rec}
Euclidean triplet	0.8711 ± 0.0073	0.8711 ± 0.0073	0.0001 ± 0.0001	0.0000 ± 0.0000	0.0001 ± 0.0001
Residual task-only	0.8769 ± 0.0070	0.8770 ± 0.0079	0.2634 ± 0.0098	0.0152 ± 0.0018	0.2801 ± 0.0103
AMN-inspired balanced	0.8749 ± 0.0056	0.8749 ± 0.0058	0.1587 ± 0.0056	0.0084 ± 0.0013	0.1524 ± 0.0064
AMN-inspired geom-first	0.8734 ± 0.0051	0.8733 ± 0.0053	0.1338 ± 0.0064	0.0068 ± 0.0012	0.1241 ± 0.0062

Table 9 reports paired differences relative to the residual task-only baseline with descriptive 95% confidence intervals. The supplementary CSV files also include paired t-tests and Wilcoxon signed-rank tests; all p-values are uncorrected and are used only to show seed-level consistency of the defect reductions, not to claim broad distribution-free

superiority across data sets. For both AMN-inspired variants, the reductions in all three finite defects are consistent across the ten seeds. The Euclidean triplet line is included only as a rigid reference.

Table 9. Paired Fashion-MNIST differences relative to residual task-only. Entries are mean paired differences with descriptive 95% confidence intervals. Negative defect differences indicate lower finite defects.

Variant	Δ Recall@1 (95% CI)	ΔD_{hom} (95% CI)	ΔD_{sub} (95% CI)	ΔD_{rec} (95% CI)
AMN-balanced	−0.0020 [−0.0063, +0.0024]	−0.105 [−0.112, −0.098]	−0.0068 [−0.0077, −0.0059]	−0.128 [−0.135, −0.120]
AMN geom-first	−0.0035 [−0.0072, +0.0003]	−0.130 [−0.137, −0.122]	−0.0084 [−0.0093, −0.0075]	−0.156 [−0.163, −0.149]
Euclidean triplet	−0.0058 [−0.0097, −0.0018]	−0.263 [−0.270, −0.256]	−0.0152 [−0.0165, −0.0139]	−0.280 [−0.287, −0.273]

Figure 6 displays the per-seed finite defects. The plots show that the reported defect reductions are not driven by a single anomalous seed: both AMN-inspired variants are below the residual task-only curve for D_{hom}, D_{sub}, and D_{rec} throughout the paired run.

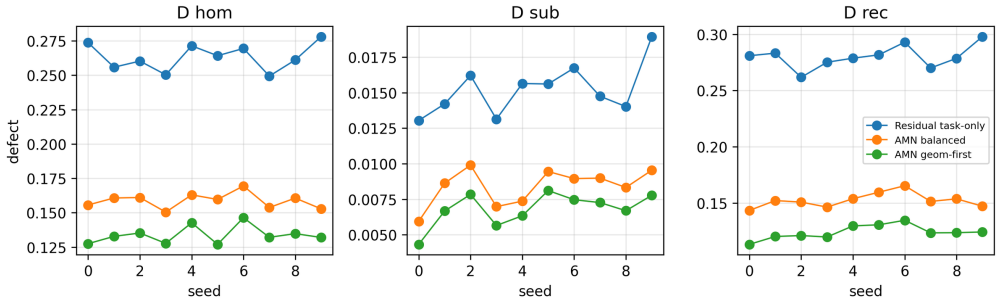


Figure 6. Per-seed Fashion-MNIST finite defects. Each point corresponds to one seed. AMN-inspired variants consistently reduce the three diagnostics relative to the residual task-only baseline.

Figures 7 and 8 provide visual summaries of the controlled Fashion-MNIST defect reductions and retrieval–geometry trade-off.

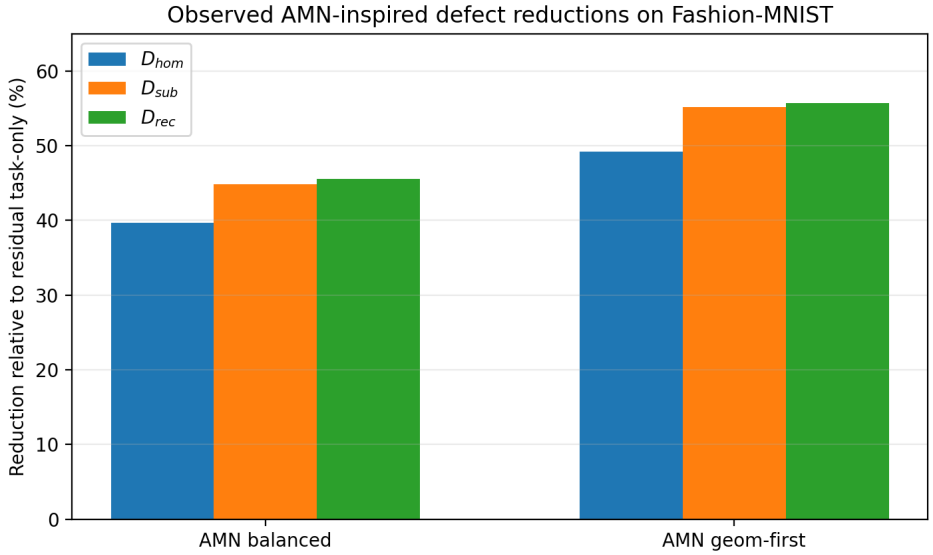


Figure 7. Fashion-MNIST reduction of finite AMN-inspired defects relative to the residual task-only baseline. Larger positive bars indicate larger reductions in D_{hom}, D_{sub}, and D_{rec}.

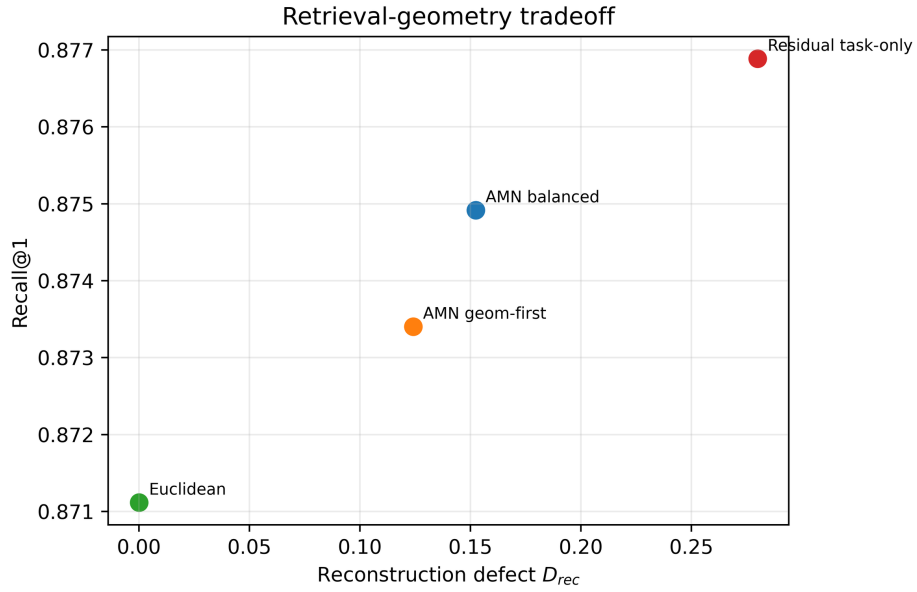


Figure 8. Fashion-MNIST trade-off between Recall@1 and reconstruction defect. The proposed variants move the residual model toward lower reconstruction defect while keeping retrieval performance close to the task-only residual baseline.

4.2. Full Official Fashion-MNIST Split Audit and Modern Baselines

The controlled subset study above is designed for ablations, robustness probes, and paired-diagnostic figures. To address full-data evaluation and modern metric-learning context under the same pre-fixed residual weights, we also ran the complete official Fashion-MNIST split with 60,000 training-gallery examples and 10,000 test queries. This full-split audit uses ten seeds and compares six variants: Euclidean triplet, supervised contrastive, proxy-anchor-style, residual task-only, AMN-balanced, and AMN-geometry-first. Reconstructed-distance retrieval is skipped in this full-split audit to keep the 10,000 × 60,000 evaluation tractable, but the finite reconstruction defect D_{rec} is still computed.

Table 10 shows two complementary facts. First, the modern Euclidean supervised-contrastive baseline obtains the highest Recall@1, 0.9099 ± 0.0017, and the proxy-anchor-style baseline gives the highest Recall@5 and Recall@10. Thus, the proposed residual family is not presented as a state-of-the-art retrieval method. Second, within the residual family, AMN-balanced and AMN-geometry-first substantially reduce finite defects relative to residual task-only. AMN-balanced changes Recall@1 from 0.8801 ± 0.0020 to 0.8758 ± 0.0043 while reducing D_{hom}, D_{sub} and D_{rec} by approximately 42.2%, 50.2% and 47.2%, respectively. AMN-geometry-first reduces the same defects by approximately 50.1%, 59.3%, and 55.8%, with a slightly higher Recall@1 cost.

Table 10. Full official Fashion-MNIST split audit. Values are the mean ± standard deviation over ten seeds. Euclidean baselines are norm-like by construction and are included to contextualize retrieval; the residual-family comparison tests the finite AMN-inspired penalties.

Variant	Recall@1	Recall@5	Recall@10	mAP@200	D_{hom}	D_{sub}	D_{rec}
Euclidean triplet	0.8698 ± 0.0065	0.9571 ± 0.0034	0.9731 ± 0.0019	0.8710 ± 0.0051	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Supervised contrastive	0.9099 ± 0.0017	0.9606 ± 0.0019	0.9712 ± 0.0015	0.9093 ± 0.0023	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Proxy-anchor style	0.8782 ± 0.0034	0.9658 ± 0.0010	0.9807 ± 0.0010	0.8586 ± 0.0028	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Residual task-only	0.8801 ± 0.0020	0.9619 ± 0.0013	0.9763 ± 0.0014	0.8778 ± 0.0026	0.3055 ± 0.0250	0.0184 ± 0.0037	0.3032 ± 0.0263
AMN-balanced	0.8758 ± 0.0043	0.9606 ± 0.0019	0.9759 ± 0.0013	0.8739 ± 0.0033	0.1766 ± 0.0179	0.0092 ± 0.0017	0.1600 ± 0.0203
AMN geom-first	0.8749 ± 0.0029	0.9611 ± 0.0016	0.9758 ± 0.0011	0.8730 ± 0.0031	0.1525 ± 0.0147	0.0075 ± 0.0013	0.1339 ± 0.0161

The same full-split audit also records the mean scale ratio

$$\bar{q} = \mathbb{E}\left[\frac{d_\theta(v)}{\|v\|_2 + \varepsilon}\right],$$

computed on sampled gallery differences. Table 11 shows that the residual models do not simply collapse to the unit Euclidean distance: their mean ratios are around 1.46–1.48, while the Euclidean baselines have $\bar{q} = 1$ by construction. At the same time, AMN-inspired training reduces the dyadic homogeneity defects at scales 2 and 4. This diagnostic does not replace a direct Euclidean-shrinkage ablation, but it documents that the learned residual head remains a nontrivial bounded deformation of the Euclidean norm under the reported protocol.

Table 11. Full-split Fashion-MNIST scale-factor audit. Here $\bar{q}$ is the mean ratio $d_\theta(v)/(\|v\|_2 + \varepsilon)$ over sampled training-gallery differences. Euclidean baselines have $\bar{q} = 1$ by construction.

Variant	$\bar{q}$	$D_{\text{hom},2}$	$D_{\text{hom},4}$	Triangle Rate
Euclidean triplet	1.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Supervised contrastive	1.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Proxy-anchor style	1.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Residual task-only	1.4567 ± 0.0127	0.2100 ± 0.0208	0.3835 ± 0.0357	0.1229 ± 0.0097
AMN-balanced	1.4714 ± 0.0165	0.1132 ± 0.0126	0.1952 ± 0.0244	0.1288 ± 0.0135
AMN geom-first	1.4848 ± 0.0145	0.0976 ± 0.0104	0.1634 ± 0.0195	0.1198 ± 0.0119

Table 12 reports seed-paired full-split residual-family effects relative to the residual task-only model.

Table 12. Paired full-split residual-family effects relative to residual task-only. Entries are mean paired differences with descriptive 95% confidence intervals; defect columns also report the relative mean reduction.

Variant	Δ Recall@1	ΔD_{hom}	ΔD_{sub}	ΔD_{rec}
AMN-balanced	−0.0044 [−0.0077, −0.0011]	−0.129 [−0.136, −0.121] (42.2%)	−0.0093 [−0.0110, −0.0075] (50.2%)	−0.143 [−0.150, −0.137] (47.2%)
AMN geom-first	−0.0052 [−0.0079, −0.0025]	−0.153 [−0.162, −0.144] (50.1%)	−0.0109 [−0.0129, −0.0090] (59.3%)	−0.169 [−0.178, −0.161] (55.8%)

Figures 9 and 10 visualize the same full-split residual-family reductions and retrieval–geometry trade-off.

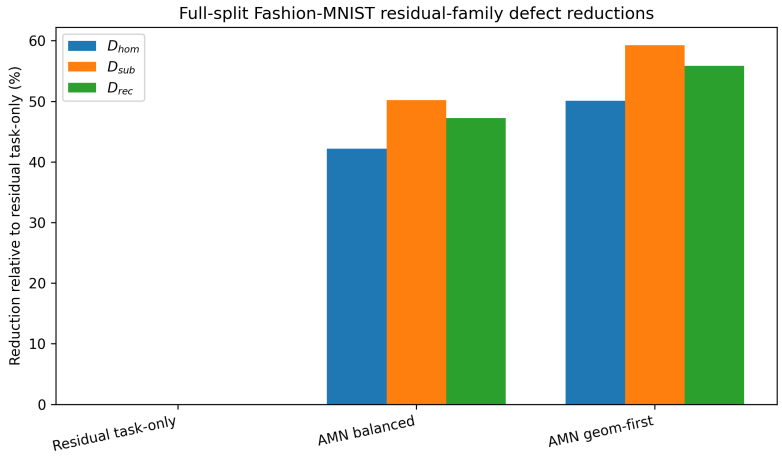


Figure 9. Full-split Fashion-MNIST residual-family defect reductions relative to residual task-only. The Euclidean modern baselines are omitted from this plot because their defects are nearly zero by construction.

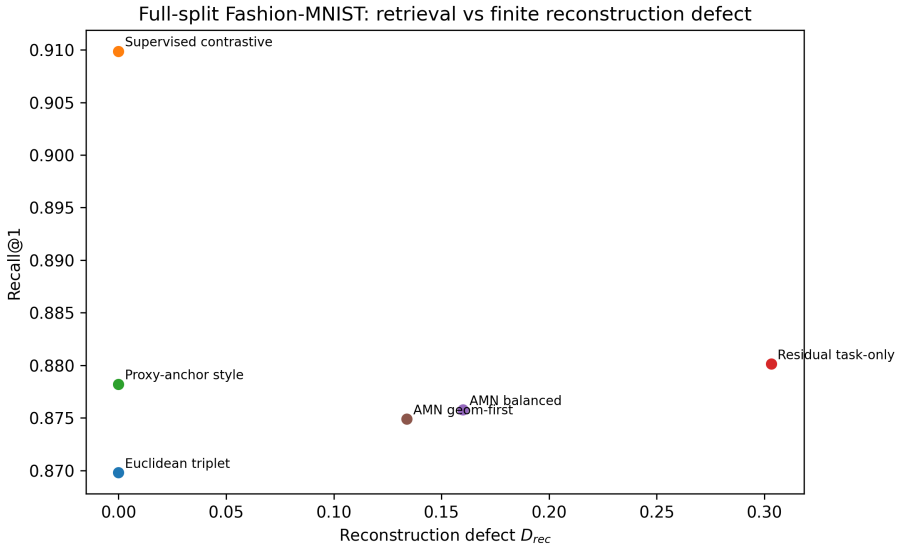


Figure 10. Full-split Fashion-MNIST retrieval–geometry trade-off. Supervised contrastive gives the strongest Recall@1 among the Euclidean baselines, while AMN-inspired residual variants move the flexible residual model toward lower reconstruction defect.

4.3. Post Hoc Finite-Geometry Audits on Fashion-MNIST

The theoretical additions in Propositions 2 and 3 and Lemma 1 motivate targeted post hoc audits that are not used for model selection. These audits were run on the official Fashion-MNIST split for three seeds. Their role is to test whether the observed finite-geometry effect survives simple alternative explanations: direct shrinkage toward Euclidean distance, changes in the residual flexibility parameter a, off-training-scale diagnostics, and extrapolation of latent operations.

A direct Euclidean-shrinkage control is the strongest cautionary baseline. Starting from the residual task-only dissimilarity, we define

$$d_\alpha(v) = (1-\alpha)d_{\mathrm{res}}(v) + \alpha\|v\|_2, \qquad \alpha \in \{0, 0.25, 0.5, 0.75, 1\}. \tag{71}$$

Table 13 shows that increasing α reduces the finite defects almost monotonically with essentially unchanged Recall@1. Thus, a substantial part of the defect reduction can also be obtained by explicitly moving the dissimilarity toward the Euclidean norm. This does not invalidate the AMN-inspired regularizer, but it sharpens its interpretation: the proposed losses provide a structured, trainable norm-like bias, whereas pure shrinkage is a post hoc collapse toward the anchor. The comparison is included to avoid attributing every finite-defect reduction uniquely to the AMN structure.

Table 13. Fashion-MNIST Euclidean-shrinkage control. The table reports post hoc mixtures $d_\alpha = (1-\alpha)d_{\mathrm{res}} + \alpha\|v\|_2$ over three seeds on the official split.

α	Recall@1	D_{hom}	D_{sub}	D_{rec}
0.00	0.8797 ± 0.0021	0.2912 ± 0.0251	0.0167 ± 0.0024	0.2842 ± 0.0216
0.25	0.8800 ± 0.0019	0.1929 ± 0.0188	0.0094 ± 0.0020	0.1617 ± 0.0180
0.50	0.8799 ± 0.0020	0.1210 ± 0.0119	0.0053 ± 0.0014	0.0844 ± 0.0077
0.75	0.8800 ± 0.0019	0.0610 ± 0.0063	0.0016 ± 0.0004	0.0361 ± 0.0031
1.00	0.8801 ± 0.0018	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000

The second audit varies the residual-scale parameter a in the head defined in Equation (4). Table 14 shows the expected behavior. When a is small, both the task-only

and AMN variants are already close to Euclidean and have small defects. As a increases, the task-only residual head becomes much less norm-like, whereas AMN-balanced controls the growth of D_{hom} and D_{rec} much more strongly. This supports the gauge interpretation: the AMN-inspired terms matter most when the residual family is flexible enough to deform the Euclidean unit-ball substantially.

Table 14. Fashion-MNIST residual-scale sweep over a. Values are means and standard deviations over three seeds. The sweep is diagnostic and is not used to select the reported full-split model.

a	Residual R@1	Residual D_{hom}	Residual D_{rec}	AMN R@1	AMN D_{hom}	AMN D_{rec}
0.25	0.8790 ± 0.0032	0.0939 ± 0.0120	0.0603 ± 0.0025	0.8722 ± 0.0046	0.0899 ± 0.0112	0.0617 ± 0.0051
0.50	0.8795 ± 0.0020	0.2055 ± 0.0175	0.1695 ± 0.0125	0.8773 ± 0.0052	0.1517 ± 0.0163	0.1282 ± 0.0132
0.65	0.8797 ± 0.0021	0.2932 ± 0.0240	0.2855 ± 0.0208	0.8802 ± 0.0011	0.1661 ± 0.0176	0.1470 ± 0.0174
0.80	0.8807 ± 0.0012	0.4478 ± 0.0292	0.5167 ± 0.0332	0.8790 ± 0.0035	0.1746 ± 0.0164	0.1558 ± 0.0156
0.90	0.8823 ± 0.0004	0.6520 ± 0.0455	0.8809 ± 0.0669	0.8788 ± 0.0053	0.1747 ± 0.0206	0.1571 ± 0.0215

The third audit evaluates diagnostics away from the exact training scale set. The training losses use $\Lambda = \{1/2, 2, 4\}$ and $L = 4$, whereas Table 15 reports held-out latent-difference defects on the test split at intermediate or larger scales, including $\sqrt{2}$, 1.5, 3, and 6, together with reconstruction at $L = 4$ and sampled subadditivity at $v + w$. The AMN-inspired variants reduce the off-grid homogeneity and reconstruction diagnostics as well. This does not prove global scale regularity, because Lemma 1 would require a Lipschitz constant in log-scale that is not certified here, but it gives empirical evidence that the effect is not restricted to the exact dyadic factors used during training.

Finally, Table 16 reports how far the latent operations used by the finite diagnostics move away from the observed-difference distribution. The reference distribution is the norm of observed test latent differences. The $v + w$ operation is moderately extrapolative, while $2v$, $4v$, and especially $16v$ are frequently outside the observed-difference 95th percentile. This confirms an important limitation: the dyadic estimator regularizes an ambient latent extension of the dissimilarity, not only on-manifold pair differences. In this sense, the finite AMN-inspired loss should be interpreted as a geometric prior on the learned latent vector space.

Table 15. Fashion-MNIST test-split off-training-scale diagnostics. H_λ is the normalized homogeneity defect at scale λ, $R_{L=4}$ is the reconstruction defect and S_{v+w} is the sampled subadditivity defect. Values are means and standard deviations over three seeds.

Variant	$H_{\sqrt{2}}$	$H_{1.5}$	H_3	H_6	$R_{L=4}$	S_{v+w}
Residual task-only	0.1029 ± 0.0118	0.1227 ± 0.0143	0.3256 ± 0.0211	0.4366 ± 0.0350	0.3044 ± 0.0247	0.0185 ± 0.0034
AMN-balanced	0.0606 ± 0.0067	0.0696 ± 0.0076	0.1600 ± 0.0166	0.2167 ± 0.0239	0.1539 ± 0.0172	0.0081 ± 0.0011
AMN geom-first	0.0526 ± 0.0050	0.0602 ± 0.0056	0.1322 ± 0.0111	0.1760 ± 0.0157	0.1259 ± 0.0114	0.0064 ± 0.0012

Table 16. Fashion-MNIST latent-operation scale audit on test latent differences. Ratios compare each operation's 95th percentile norm with the observed-difference 95th percentile; "above" is the fraction exceeding that observed 95th percentile.

Variant	$v + w$ Ratio	$v + w$ Above	$2v$ Ratio	$2v$ Above	$4v$ Ratio	$4v$ Above	$16v$ Ratio	$16v$ Above
Residual task-only	1.48	25.2%	2.00	57.2%	4.01	86.4%	16.03	97.7%
AMN-balanced	1.48	25.4%	2.00	57.5%	4.00	85.9%	16.01	97.1%
AMN geom-first	1.48	25.6%	2.00	57.9%	4.00	85.7%	16.02	97.0%

4.4. Secondary Non-Visual UCI HAR Corroboration

Table 17 reports the non-visual UCI HAR corroboration after extending the CPU run to ten seeds and full official evaluation. The subject-based train/test split is fixed; the seeds affect initialization and minibatch sampling. All 2947 test records are used as queries, and all 7352 training records form the gallery. The result remains secondary because the architecture and weights were not separately tuned for UCI HAR, but it is no longer only a small three-seed probe.

Within the residual family, AMN-inspired balanced reduces D_{hom} from 0.3637 ± 0.0364 to 0.0547 ± 0.0169, D_{sub} from 0.0368 ± 0.0160 to 0.0067 ± 0.0019, and D_{rec} from 0.4501 ± 0.0545 to 0.0583 ± 0.0214. These correspond to paired reductions of approximately 85.0%, 81.9%, and 87.0%, while Recall@1 changes by only -0.0022 ± 0.0094. The geometry-first variant behaves similarly. The Euclidean triplet baseline obtains the highest absolute Recall@1 and is almost defect-free because it is a norm by construction; the residual-family conclusion is therefore geometric rather than a claim of best retrieval performance.

Table 17. UCI HAR non-visual corroboration with full-query/full-gallery evaluation. Values are mean ± standard deviation over ten seeds on the fixed official subject split.

Variant	Recall@1	Recall@5	D_{hom}	D_{sub}	D_{rec}
Euclidean triplet	0.9480 ± 0.0038	0.9767 ± 0.0025	0.0001 ± 0.0001	0.0000 ± 0.0000	0.0001 ± 0.0001
Residual task-only	0.9234 ± 0.0068	0.9696 ± 0.0046	0.3637 ± 0.0364	0.0368 ± 0.0160	0.4501 ± 0.0545
AMN-balanced	0.9212 ± 0.0080	0.9690 ± 0.0036	0.0547 ± 0.0169	0.0067 ± 0.0019	0.0583 ± 0.0214
AMN geom-first	0.9221 ± 0.0088	0.9700 ± 0.0048	0.0505 ± 0.0156	0.0065 ± 0.0019	0.0577 ± 0.0195

Table 18 reports the corresponding seed-paired UCI HAR effects.

Table 18. Paired UCI HAR effects relative to residual task-only. Reductions are computed from seed-paired differences on the fixed official split.

Variant	ΔRecall@1	Reduction D_{hom}	Reduction D_{sub}	Reduction D_{rec}
AMN-balanced	-0.0022 ± 0.0094	85.0%	81.9%	87.0%
AMN geom-first	-0.0013 ± 0.0104	86.1%	82.3%	87.2%
Euclidean triplet	0.0246 ± 0.0068	100.0%	100.0%	100.0%

Figure 11 shows the seed-level UCI HAR finite-defect profiles.

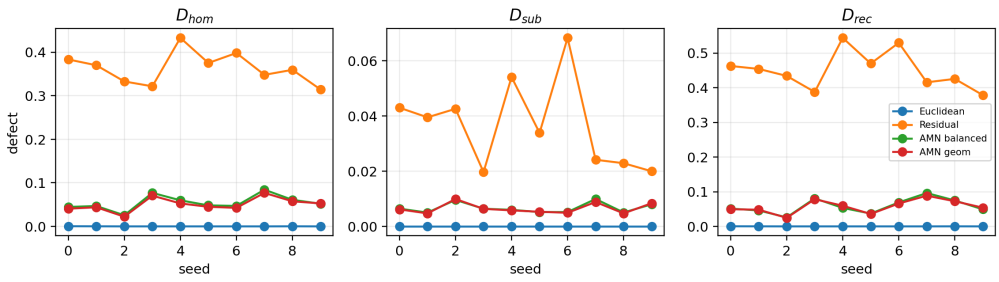


Figure 11. Per-seed UCI HAR finite-defect summary over ten seeds.

The earlier UCI HAR latent-operation diagnostic remains useful for interpretation: operations such as $v + w$, $2v$, and $4v$ probe the ambient latent-vector space rather than only observed pair differences. The new Fashion-MNIST latent-scale audit in Table 16 shows the same limitation on the primary image benchmark and therefore makes this point part of the main empirical interpretation.

4.5. Auxiliary Digits Image-Domain Check

The scikit-learn Digits run is an auxiliary image-domain mechanism check and is not part of the core empirical claim. Its summary is moved to Appendix C to keep the main results focused on Fashion-MNIST, UCI HAR, and the diagnostic audits. The qualitative observation is that reducing finite-scale defects is accompanied by a more stable dyadic reconstruction on this smaller dataset as well.

4.6. Non-AMN Controls

A central question is whether the observed effect is merely generic regularization. Table 19 gives evidence against that interpretation. Stronger weight decay leaves the AMN-inspired defects essentially unchanged. The smoothness control worsens the joint defect profile. Reconstruction-only regularization gives modest improvements. Homogeneity plus subadditivity without the explicit reconstruction term is much stronger, reducing D_{rec} by 43.2%; this supports the finite-stability statement, because reconstruction improves when normalized slopes and subadditive behavior are controlled. The full AMN-inspired variants provide the best joint defect reductions. Figure 12 summarizes the reduction profile of the AMN-specific and generic regularization controls.

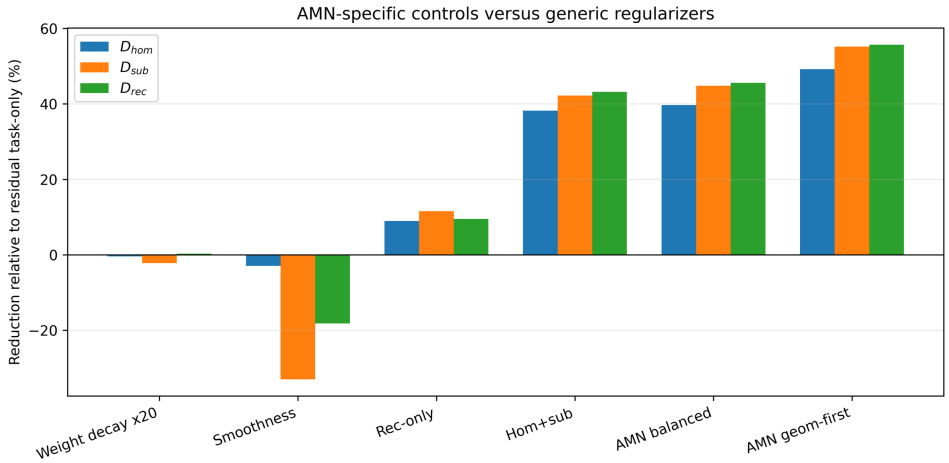


Figure 12. AMN-specific controls versus generic regularizers. Reductions are measured relative to the residual task-only baseline. Generic regularization does not reproduce the joint reduction in D_{hom}, D_{sub}, and D_{rec}.

Table 19. Regularizer controls on Fashion-MNIST. Values are the mean ± standard deviation over ten seeds. Reductions are relative to the residual task-only dissimilarity; negative reductions indicate worse defect values.

Variant	Recall@1	D_{hom}	D_{sub}	D_{rec}	Red. D_{hom}	Red. D_{sub}	Red. D_{rec}
Residual task-only	0.8769 ± 0.0070	0.2634 ± 0.0098	0.0152 ± 0.0018	0.2801 ± 0.0103	0.0%	0.0%	0.0%
Weight decay ×20	0.8775 ± 0.0050	0.2645 ± 0.0096	0.0156 ± 0.0020	0.2791 ± 0.0095	−0.4%	−2.2%	0.3%
Smoothness	0.8758 ± 0.0048	0.2711 ± 0.0464	0.0202 ± 0.0116	0.3309 ± 0.1051	−2.9%	−33.0%	−18.2%
Reconstruction-only	0.8757 ± 0.0062	0.2397 ± 0.0098	0.0135 ± 0.0017	0.2536 ± 0.0086	9.0%	11.6%	9.5%
Homogeneity + subadditivity	0.8733 ± 0.0046	0.1628 ± 0.0053	0.0088 ± 0.0013	0.1590 ± 0.0065	38.2%	42.2%	43.2%
AMN-inspired balanced	0.8749 ± 0.0056	0.1587 ± 0.0056	0.0084 ± 0.0013	0.1524 ± 0.0064	39.7%	44.8%	45.6%
AMN-inspired geom-first	0.8734 ± 0.0051	0.1338 ± 0.0064	0.0068 ± 0.0012	0.1241 ± 0.0062	49.2%	55.2%	55.7%

4.7. Robustness Probes

Robustness probes are secondary and mixed. As reported in Appendix C, AMN-inspired variants improve the Gaussian-noise probe relative to residual task-only, are close on shift, and are slightly worse on occlusion and contrast. These diagnostics suggest that the regularizer changes geometry, but they are not used to claim general robustness.

5. Discussion

The results support a geometric interpretation rather than a pure performance interpretation. In the controlled Fashion-MNIST subset study, the residual task-only dissimilarity obtains the best Recall@1 among the flexible residual models, but it has large homogeneity and reconstruction defects. AMN-inspired balanced sacrifices only about 0.002 absolute Recall@1 while reducing all three finite defects by approximately 40–46%, and AMN-inspired geom-first reduces the defects by approximately 49–56% with a still modest retrieval cost. The full official Fashion-MNIST audit repeats the residual-family comparison on all 60,000/10,000 examples: AMN-balanced reduces D_{hom}, D_{sub} and D_{rec} by approximately 42%, 50% and 47% relative to residual task-only, while AMN-geometry-first reduces them by approximately 50%, 59% and 56%. The same full-split run also shows that modern Euclidean objectives, especially supervised contrastive learning, can achieve higher Recall@1 than the residual family. This is consistent with the stated scope: the AMN-inspired terms are designed to control finite geometry inside a flexible residual family, not to maximize retrieval accuracy across all metric-learning objectives. The non-visual UCI HAR experiment now uses ten seeds and the full official query/gallery evaluation. It points in the same geometric direction on tabular sensor-derived data: relative to the residual task-only dissimilarity, AMN-inspired balanced reduces D_{hom}, D_{sub}, and D_{rec} by approximately 85%, 82% and 87%, while leaving Recall@1 within about 0.002 of the residual task-only mean.

5.1. Kernel-Derived Distance Quotient Interpretation

The clearest way to connect the finite construction with the AMN intuition is through the unbounded distance-like quantities encoded by learned similarities. A bounded kernel value, or the bounded RKHS distance induced by a normalized kernel, has trivial large-scale AMN geometry. The object that can carry AMN information is obtained from the exponent: for Laplacian forms, it is the negative-log distance $-\tau \log K$, while for Gaussian forms, the negative log is squared-distance-like, and the distance-like object is its square root. In the model case of Proposition 4, a learned kernel-derived distance satisfying $D(x,y) = \|A(x-y)\| + O(1)$ has stable slope $\nu(v) = \|Av\|$ and bounded-direction space $E_0 = \ker A$. Thus, the quotient E/E_0 is the effective semantic latent space, and ν is the stable semantic norm on that quotient. This is the large-scale ideal that the finite norm-envelope and refinement-tower diagnostics approximate: the finite witness norms p_S and their refinements are sampled substitutes for the quotient norm recovered from a noisy learned distance.

The energy terminology should therefore be read as a stability principle for geometric costs rather than as a claim of physical energy conservation. What is stable is the asymptotic specific energy $D(nv,0)/n$: by Corollary 2, bounded local perturbations do not change its limit. This is exactly why the quotient interpretation is useful for learned embeddings with noise, quantization, or saturation. The same corollary also explains the kernel side of the construction: if the energy is subadditive, then the exponential affinity is multiplicatively path-consistent, $K(x,z) \geq K(x,y)K(y,z)$. Thus, the subadditivity defect can be interpreted either geometrically, as failure of a norm-like triangle inequality, or in similarity language, as failure of multiplicative consistency along paths.

The entropy analogy is also useful but limited. A raw value $-\log K(x,y)$ is a logarithmic cost, not an entropy. Once the affinities are normalized into a conditional neighborhood distribution $p(y \mid x)$, the quantity $-\log p(y \mid x)$ is a surprisal term, and its expectation is an entropy. Hence, AMN-inspired regularization acts on geometric energy, whereas entropy describes uncertainty of the probability distribution induced by that energy. This distinction prevents a common confusion: stabilizing the log-kernel distance is not the same as minimizing an entropy, although both use logarithmic quantities.

This interpretation also explains why the residual family is anchored but still meaningful. The Euclidean factor supplies a reference norm and prevents pathological signs or degeneracies; the residual head modifies the radial factor of the learned distance-like quantity. AMN-inspired penalties then ask whether this modified dissimilarity is stable along rays, approximately convex through subadditivity, and reconstructible by dyadic slopes. In this sense, the regularizer is not merely a penalty on three diagnostics; it is a finite attempt to recover a stable quotient norm-like distance from a noisy latent dissimilarity.

5.2. Effect of Euclidean Anchoring

A central limitation is that the residual family is not an arbitrary learned dissimilarity. By construction, $\rho_\theta(v) - \delta$ is a bounded multiplicative deformation of $\|v\|_2$. For $a = 0.65$, the deformation factor lies between 0.35 and 1.65. Thus, the experiments do not show that the method recovers normability from an unconstrained black-box distance. They show that, inside a Euclidean-anchored residual family with useful flexibility, the finite AMN-inspired terms reduce scale, subadditivity, and reconstruction defects. This anchoring is a deliberate design choice: it prevents negative or wildly unbounded scores and gives a stable reference for retrieval, but it narrows the scope of the AMN interpretation.

The new post hoc audits make this limitation explicit. The residual-factor table shows that the trained residual heads do not collapse exactly to the unit Euclidean norm: their scale factors have substantial spread and 5–95% ranges far from a point mass at 1. The residual-scale sweep further shows that, as the flexibility parameter a grows, the task-only residual head becomes much less norm-like, while AMN-balanced controls the growth of D_{hom} and D_{rec} much more strongly. At the same time, the direct Euclidean-shrinkage control is an important caution. Mixing the task-only residual dissimilarity with $\|v\|_2$ reduces finite defects almost monotonically and can do so with essentially unchanged Recall@1. Consequently, the AMN-inspired loss should not be described as the only way to reduce these diagnostics. Its value is that it provides a differentiable, structured gauge-oriented bias during training, whereas shrinkage is an explicit post hoc contraction toward the anchor.

The role of the Euclidean baselines follows from this limitation. Euclidean triplet, supervised contrastive, and proxy-anchor-style objectives are almost defect-free under D_{hom}, D_{sub}, and D_{rec} because they use a norm by construction. On UCI HAR, the Euclidean triplet baseline also has the best absolute Recall@1, and on Fashion-MNIST, supervised contrastive is best at Recall@1. Therefore, the method should not be judged by whether it beats Euclidean geometry on every performance metric. The targeted question is whether a flexible residual geometry can be made more norm-like and auditable. The paired Fashion-MNIST statistics, the a-sweep, and the UCI HAR corroboration indicate that it can, at least for the finite sampled scales and architectures tested here.

5.3. AMN Interpretation, Off-Grid Scales and Latent Extrapolation

The synthetic example clarifies the finite mathematics before considering data: exact norms are reconstructed exactly by the dyadic estimator, while scale-perturbed non-norms exhibit nonzero homogeneity and reconstruction defects. The homogeneous nonconvex gauge shows that homogeneity and reconstruction do not imply subadditivity, and the dyadic-aliasing perturbation shows why off-grid scale checks are useful. The learned Fashion-MNIST off-grid audit follows this lesson empirically. AMN-inspired variants reduce homogeneity defects at scales such as $\sqrt{2}$, 1.5, 3, and 6, even though not all these factors are used in training. This is consistent with the log-scale regularity motivation in Lemma 1, but it is still an empirical audit rather than a certified global guarantee.

The latent-operation scale diagnostics sharpen another caveat. Operations such as $v + w$, $2v$, $4v$, and $16v$ do not necessarily correspond to differences between observed data points. On Fashion-MNIST, $v + w$ is moderately extrapolative, whereas $2v$, $4v$, and especially $16v$ frequently exceed the observed-difference 95th percentile. Since $\widehat{N}_L$ with $L = 4$ evaluates dyadic multiples up to $16v$, the estimator should be understood as a regularizer of the ambient latent extension, not as an on-manifold certificate. This does not make the regularizer invalid: regularizing an ambient extension can stabilize a learned function. However, it means that the method should not be interpreted as certifying the geometry of the data manifold itself.

The control experiments then clarify the learned mechanism. Weight decay changes capacity but does not repair finite AMN-inspired defects. Smoothness penalizes gradients but does not enforce scaling or subadditivity. Reconstruction-only regularization improves D_{rec} mildly, but the homogeneity-plus-subadditivity ablation improves D_{rec} substantially even without directly optimizing it. This is consistent with Propositions 1 and 2: reconstruction becomes stable when normalized dyadic slopes vary less and subadditive behavior is controlled.

The finite norm-envelope construction strengthens this interpretation. Given a sampled set of latent differences S, the convex hull of normalized atoms $\pm v/\rho(v)$ defines an explicit witness norm p_S. The envelope gap $\Gamma_S(\rho)$ measures how much of the sampled dissimilarity is lost when the sampled unit ball is convexified. Thus, the subadditivity audit can be read geometrically: a small sampled convexification gap means that the learned dissimilarity is close, on S, to a concrete norm rather than only to a collection of pointwise inequalities. The dyadic-to-envelope bound then separates two finite errors: slope instability along sampled rays and convexification gap across sampled directions. This does not certify a global norm, but it gives a sharper finite target for future audits and for possible training losses based directly on p_S.

The refinement-tower formulation makes the connection with AMN closer and gives a finite version of the kernel-derived distance quotient picture. A single finite witness norm p_S is only a sampled certificate. A sequence of refined certificates, with more directions, more scales, or improved residual heads, can be interpreted as a finite AMN tower when consecutive witness norms are coherent. Under the hypotheses of Theorem 4, these finite witnesses converge locally uniformly to a limiting norm. If the lower anchoring degenerates, the same construction yields a seminorm and hence a quotient norm after factoring out collapsed directions. This provides a mathematical way to describe successive refinement of learned latent spaces without claiming that the experiments themselves prove asymptotic AMN rigidity.

The expectation/probability and concentration corollaries further explain why small squared minibatch losses and finite Monte Carlo diagnostics should be read as control of sampled defects rather than as pointwise guarantees.

5.4. Empirical Scope

Several experimental limitations remain. Fashion-MNIST is still the only large image benchmark in the paper. The controlled subset protocol is retained because it supports the largest ablation, robustness, and paired-diagnostic suite; the full official split audit is narrower, omits reconstructed-distance retrieval for tractability, and documents full-data behavior with modern baselines. The additional Fashion-MNIST audits use three seeds and therefore function as targeted stress tests, not as a second full ten-seed benchmark. UCI HAR is a stronger non-visual corroboration than before because it now uses ten seeds and all official queries/gallery records, but it still uses a compact MLP encoder and weights

transferred from the exploratory protocol rather than a separate UCI HAR search. Digits are retained only as a small auxiliary image-domain check.

The full-split modern baselines clarify the retrieval context. Supervised contrastive learning gives the best Recall@1 among the tested Fashion-MNIST objectives, and the proxy-anchor-style baseline gives strong Recall@5/10. However, these baselines are Euclidean embeddings and are therefore almost defect-free under D_{hom}, D_{sub}, and D_{rec} by construction; they do not test whether finite AMN-inspired penalties repair a flexible residual dissimilarity. Accordingly, the manuscript should be read as a finite geometric regularization and diagnostic study, complemented by modern retrieval baselines, not as a broad claim of competitive metric-learning performance.

The proposed diagnostics also have failure modes. They may be small for a Euclidean embedding that is strong geometrically but not optimal for a task, or large for a task-adapted dissimilarity that improves ranking while deliberately breaking norm-like structure. Dyadic tests can miss oscillations at intermediate scales, which is why off-grid audits are included. Subadditivity probes on $v + w$ can leave the observed data-difference manifold, and finite samples may miss rare directions with large defects. Low values of D_{hom}, D_{sub}, and D_{rec} should therefore be interpreted as evidence of improved sampled latent norm-likeness, not as universal proof of useful semantic geometry.

From a computational point of view, the minibatch losses scale with sampled latent differences and can be applied to larger encoders, including pretrained embeddings or transformer representations, by placing the residual head on top of fixed or fine-tuned latent vectors. The most expensive object is the finite norm-envelope witness p_S, which is a linear-programming diagnostic on a chosen atom set and is therefore better suited to small audits or compressed representative sets than to every training step. Scaling the envelope computation by subsampling, coresets, or learned norm families is a natural next step.

Finally, the AMN connection is motivational and finite: the paper does not prove AMN rigidity, a finite necessary-and-sufficient normability criterion, or a global metric certificate. Robustness probes are exploratory and mixed. Larger image retrieval data sets beyond Fashion-MNIST, raw sensor sequences, text embeddings, graph-distance settings, less Euclidean-anchored residual heads, learned norm families, scalable finite-envelope approximations, and AMN-inspired versions of modern objectives are natural next benchmarks.

6. Conclusions

This paper introduced a finite AMN-inspired framework for diagnosing and regularizing norm-like geometry in Euclidean-anchored residual neural dissimilarities. The central object is not a global input-space metric but a finite latent dissimilarity evaluated on sampled differences. The proposed diagnostics measure homogeneity, subadditivity, and dyadic reconstruction defects; the associated theory explains these quantities through finite dyadic stability, learned gauges, finite norm-envelope witnesses, refinement towers, and quotient behavior of stable kernel-derived distances.

The experiments support the intended geometric claim. On the full official Fashion-MNIST split, supervised-contrastive and proxy-style Euclidean objectives provide strong retrieval baselines and can outperform the residual family in Recall@1. Inside the residual family, however, AMN-inspired regularization reduces finite norm-like defects substantially relative to the task-only residual head. UCI HAR gives the same qualitative signal on non-visual sensor-derived features. The shrinkage, residual-scale, off-grid, and latent-extrapolation audits further clarify the mechanism and its limitations: part of the effect is related to the Euclidean anchor, off-grid defects are still empirical rather than certi-

fied, and dyadic operations may regularize ambient latent regions outside the observed difference distribution.

The practical contribution is therefore a way to audit and control finite geometric structure in learned dissimilarities, complementary to retrieval accuracy. The method is useful when one wants a flexible dissimilarity to remain interpretable as a norm-like latent cost. It should not be read as a state-of-the-art retrieval objective, a global metric certificate, or a finite proof of AMN normability. Future work should test the same diagnostics on larger pretrained embeddings, text and graph representations, fewer Euclidean-anchored heads, scalable norm-envelope approximations, and downstream tasks where quotient norm-like geometry improves transfer, calibration, or search.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math14132420/s1, The supplementary material accompanying this article is supplied as a ZIP archive. It contains software source code, configuration files, compact processed splits for the secondary data sets, final CSV result files, generated diagnostic figures, per-seed statistical summaries, the synthetic dyadic diagnostic example, the finite norm-envelope and refinement-tower CPU diagnostics, UCI HAR full-query/full-gallery summaries, Fashion-MNIST shrinkage, residual-scale, off-grid and latent-operation audits, a CPU smoke test, scripts for regenerating Fashion-MNIST, UCI HAR, and scikit-learn Digits splits, and scripts for rerunning the reported experiments and audits.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Proofs of the Finite Statements

Appendix A.1. Proof of Proposition 1

Let $\bar{\rho}(v) = (\rho(v) + \rho(-v))/2$, $a_k(v) = 2^{-k}\bar{\rho}(2^k v)$ and $\widehat{N}_L(v) = (L+1)^{-1}\sum_{k=0}^{L} a_k(v)$. For part (i),

$$\begin{aligned}\widehat{N}_L(2v) &= \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k}\bar{\rho}(2^{k+1}v)\\ &= \frac{2}{L+1}\sum_{j=1}^{L+1} 2^{-j}\bar{\rho}(2^j v) = \frac{2}{L+1}\sum_{j=1}^{L+1} a_j(v).\end{aligned}$$

Also,

$$2\widehat{N}_L(v) = \frac{2}{L+1}\sum_{j=0}^{L} a_j(v).$$

Subtracting gives

$$\widehat{N}_L(2v) - 2\widehat{N}_L(v) = \frac{2}{L+1}\big(a_{L+1}(v) - a_0(v)\big).$$

Thus, when the normalized slopes $a_k(v)$ are approximately constant, the reconstruction is approximately homogeneous.

For part (ii), multiply the assumed approximate subadditivity inequality for $\bar{\rho}$ by 2^{-k} and average:

$$\begin{aligned}\widehat{N}_L(v+w) &= \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k}\bar{\rho}(2^k(v+w)) \\ &\le \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k}\bar{\rho}(2^k v) + \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k}\bar{\rho}(2^k w) \\ &\quad + \varepsilon_s \frac{1}{L+1}\sum_{k=0}^{L} 2^{-k}\big(1+\bar{\rho}(2^k v)+\bar{\rho}(2^k w)\big) \\ &= (1+\varepsilon_s)\big(\widehat{N}_L(v)+\widehat{N}_L(w)\big) + \varepsilon_s c_L,\end{aligned}$$

where $c_L = (L+1)^{-1}\sum_{k=0}^{L} 2^{-k} \le 2/(L+1)$.

For part (iii), use the triangle inequality:

$$|\widehat{N}_L(v) - \bar{\rho}(v)| = \left|\frac{1}{L+1}\sum_{k=0}^{L}(a_k(v)-a_0(v))\right| \le \frac{1}{L+1}\sum_{k=0}^{L}|a_k(v)-a_0(v)|.$$

This proves the proposition.

Appendix A.2. Proof of Proposition 2

For any $k \ge 1$, telescoping the log-ratio bound gives

$$|\log a_k(v) - \log a_0(v)| \le \sum_{j=0}^{k-1} |\log a_{j+1}(v) - \log a_j(v)| \le k\eta.$$

Exponentiating yields Equation (24). Since $a_0(v) = \bar{\rho}(v)$,

$$\begin{aligned}\frac{|\widehat{N}_L(v)-\bar{\rho}(v)|}{\bar{\rho}(v)} &= \left|\frac{1}{L+1}\sum_{k=0}^{L}\left(\frac{a_k(v)}{a_0(v)}-1\right)\right| \\ &\le \frac{1}{L+1}\sum_{k=0}^{L}\left|\frac{a_k(v)}{a_0(v)}-1\right| \le \frac{1}{L+1}\sum_{k=0}^{L}(e^{k\eta}-1).\end{aligned}$$

The expansion follows from $e^{k\eta} = 1 + k\eta + O(k^2\eta^2)$ and $(L+1)^{-1}\sum_{k=0}^{L} k = L/2$.

Appendix A.3. Proof of Proposition 3

Evenness follows from evenness of q, and homogeneity follows because $q(\lambda v) = q(v)$ for $\lambda > 0$ and $q(-v) = q(v)$ for negative scalars. If $0 < c \le q(v) \le C < \infty$, then C^{-1} times the Euclidean unit ball is contained in B_ρ and B_ρ is contained in c^{-1} times the Euclidean unit ball. Thus, B_ρ is bounded and absorbing; central symmetry follows from evenness.

For a non-negative, positive homogeneous function, subadditivity implies convexity of the unit ball: if $x, y \in B_\rho$ and $t \in [0,1]$, then

$$\rho(tx + (1-t)y) \le t\rho(x) + (1-t)\rho(y) \le 1.$$

Conversely, if B_ρ is convex, then for $x, y \neq 0$ the points $x/\rho(x)$ and $y/\rho(y)$ belong to B_ρ. Convexity gives

$$\frac{x+y}{\rho(x)+\rho(y)} = \frac{\rho(x)}{\rho(x)+\rho(y)}\frac{x}{\rho(x)} + \frac{\rho(y)}{\rho(x)+\rho(y)}\frac{y}{\rho(y)} \in B_\rho,$$

so by homogeneity $\rho(x+y) \leq \rho(x)+\rho(y)$. The cases $x = 0$ or $y = 0$ are immediate. Positive definiteness is inherited from $q > 0$ away from zero and $\rho(0) = 0$. Hence convexity of B_ρ is equivalent to subadditivity, and in that case ρ is a norm.

Appendix A.4. Proof of Theorem 1

Since S spans E_S and is finite, the convex hull $B_S(\rho)$ is compact, centrally symmetric, and has nonempty relative interior in E_S. It is also absorbing in E_S: every $x \in E_S$ is a finite linear combination of elements of S, hence belongs to $tB_S(\rho)$ for some sufficiently large t. Because $B_S(\rho)$ is bounded and centrally symmetric and contains no line, its Minkowski functional is finite, positive away from zero, homogeneous, and subadditive. Thus, p_S is a norm on E_S.

For (i), if $u \in S$, then $u/\rho(u) \in B_S(\rho)$, and hence $u \in \rho(u)B_S(\rho)$. Therefore $p_S(u) \leq \rho(u)$.

For the representation in Equation (29), write $x \in tB_S(\rho)$ as a convex combination of atoms $\pm u_j/\rho(u_j)$. Multiplying the convex coefficients by t gives a representation $x = \sum_j \alpha_j u_j$ with cost $\sum_j |\alpha_j|\rho(u_j) \leq t$. Taking the infimum over t gives one inequality. Conversely, any representation $x = \sum_j \alpha_j u_j$ implies that x/T belongs to $B_S(\rho)$ with $T = \sum_j |\alpha_j|\rho(u_j)$, provided $T > 0$. Taking infima gives the reverse inequality.

For (ii), let p be a seminorm with $p(u) \leq \rho(u)$ for all $u \in S$. Then $p(\pm u/\rho(u)) \leq 1$ for every atom. Since the unit ball of p is convex and centrally symmetric, it contains $B_S(\rho)$. Therefore, if $x \in tB_S(\rho)$, then $p(x) \leq t$. Taking the infimum over t gives $p(x) \leq p_S(x)$ for all $x \in E_S$.

Part (iii) is the definition of $\Gamma_S(\rho)$ combined with $p_S(u) \leq \rho(u)$.

For (iv), apply Equation (32) to every representation $u = \sum_j \alpha_j u_j$ with $u, u_j \in S$. The equivalent representation formula for p_S gives

$$p_S(u) = \inf \sum_j |\alpha_j|\rho(u_j) \geq (1-\eta)\rho(u).$$

Together with $p_S(u) \leq \rho(u)$, this implies $\Gamma_S(\rho) \leq \eta$.

Appendix A.5. Proof of Corollary 1

By the triangle inequality,

$$|\widehat{N}_L(v) - p_S(v)| \leq |\widehat{N}_L(v) - \rho(v)| + |\rho(v) - p_S(v)|.$$

Dividing by $\rho(v)$, using Proposition 2 for the first term and using the definition of $\Gamma_S(\rho)$ for the second term gives Equation (33).

Appendix A.6. Proof of Theorem 2

The uniform equivalence Equation (34) implies that the restrictions of p_n to the Euclidean unit sphere are uniformly bounded above by C and uniformly Lipschitz with constant C, because every norm satisfies $|p_n(x) - p_n(y)| \leq p_n(x-y) \leq C\|x-y\|_2$. By Arzela–Ascoli, a subsequence converges uniformly on the unit sphere to a continuous function p. Extending by positive homogeneity gives locally uniform convergence on E. The limit is non-negative, even, homogeneous, and subadditive as a locally uniform limit of norms. The lower bound $p(x) \geq c\|x\|_2$ shows that $p(x) = 0$ only at $x = 0$, so p is a norm.

Let (n_j) be a subsequence such that $p_{n_j} \to p$ locally uniformly. Fix a compact set K containing the sampled sequence u_{n_j}. For every $u \in S_{n_j}$,

$$0 \le \frac{\rho_{n_j}(u) - p_{n_j}(u)}{\rho_{n_j}(u)} \le \Gamma_{S_{n_j}}(\rho_{n_j}).$$

Therefore,

$$\frac{|\rho_{n_j}(u_{n_j}) - p(u_{n_j})|}{\rho_{n_j}(u_{n_j})} \le \Gamma_{S_{n_j}}(\rho_{n_j}) + \frac{\sup_{x\in K} |p_{n_j}(x) - p(x)|}{\rho_{n_j}(u_{n_j})}.$$

The first term tends to zero by assumption, and the second tends to zero by local uniform convergence and the lower bound on $\rho_{n_j}(u_{n_j})$. This proves Equation (35).

Appendix A.7. Proof of Theorem 3

The inclusion $S_n \subset S_{n+1}$ gives $B_n \subseteq B_{n+1}$, and by definition $B_n \subseteq B_\infty$. Since the gauge of a larger unit ball is smaller, $p_{n+1} \le p_n$ on E.

It remains to justify Hausdorff and gauge convergence. Let $A_\infty = \{\pm v/\rho(v) : v \in S_\infty\}$ and $B_\infty = \overline{\text{conv}} A_\infty$. Because B_∞ is compact and $\bigcup_n B_n$ is dense in B_∞ by construction, for every $\delta > 0$ there exist finitely many points $b_1, \dots, b_M \in \bigcup_n B_n$ whose δ-balls cover B_∞. Since the sets B_n are increasing, all b_i belong to one common B_N. Therefore, $B_\infty \subseteq B_N + \delta B_2$, where B_2 denotes the Euclidean unit ball, while $B_N \subseteq B_\infty$. This proves $d_H(B_n, B_\infty) \to 0$.

Because B_∞ is a convex body, there are $0 < r \le R < \infty$ such that $rB_2 \subseteq B_1 \subseteq B_n \subseteq B_\infty \subseteq RB_2$. Hausdorff convergence of convex bodies with a common interior ball implies local uniform convergence of the gauges. For completeness, fix $\delta \in (0, r/2)$. For all large n, $B_\infty \subseteq B_n + \delta B_2$. Since B_n is convex and contains rB_2, this inclusion implies

$$B_\infty \subseteq \left(1 + \frac{\delta}{r}\right) B_n.$$

Therefore

$$\left(1 + \frac{\delta}{r}\right)^{-1} p_n(x) \le p_\infty(x) \le p_n(x),$$

for every $x \in E$. On a compact set K, the norms p_n are uniformly bounded, so $\sup_{x\in K} |p_n(x) - p_\infty(x)| \to 0$.

If $u_n \in S_n$, then by the definition of the envelope gap,

$$0 \le \rho(u_n) - p_n(u_n) \le \Gamma_{S_n}(\rho|_{S_n})\rho(u_n).$$

Combining this with the locally uniform convergence $p_n \to p_\infty$ on a compact containing u_n and the assumption $\inf_n \rho(u_n) > 0$ yields Equation (42).

Appendix A.8. Proof of Theorem 4

Let $K \subset E$ be compact. The summability assumption gives, for $m > n$,

$$\sup_{x\in K} |p_m(x) - p_n(x)| \le \sum_{j=n}^{m-1} \Xi_j(K) \to 0.$$

Thus (p_n) is a Cauchy sequence in the uniform norm on every compact set. Hence, it converges locally uniformly to a continuous function p_∞.

The uniform anchoring bounds pass to the limit, so

$$c\|x\|_2 \le p_\infty(x) \le C\|x\|_2.$$

Moreover, each p_n is even, positively homogeneous and subadditive. These three properties are preserved under locally uniform convergence, and the lower bound implies positive definiteness. Therefore, p_∞ is a norm on E.

For $u_n \in S_n$, the envelope inequality gives

$$0 \le \rho_n(u_n) - p_n(u_n) \le \Gamma_n \rho_n(u_n).$$

Let K be a compact set containing the sequence u_n. Then

$$\frac{|\rho_n(u_n) - p_\infty(u_n)|}{\rho_n(u_n)} \le \Gamma_n + \frac{\sup_{x\in K} |p_n(x) - p_\infty(x)|}{\rho_n(u_n)} \to 0.$$

Finally,

$$\frac{|\widehat{N}^{(n)}_{L_n}(u_n) - p_\infty(u_n)|}{\rho_n(u_n)} \le \Theta_n(u_n) + \frac{|\rho_n(u_n) - p_\infty(u_n)|}{\rho_n(u_n)},$$

and the right-hand side tends to zero by the previous conclusion and Equation (45). This proves Equation (46).

Appendix A.9. Proof of Proposition 4

For every $v \in E$ and every integer $n \ge 1$, Equation (48) gives

$$n\|Av\|_H - B \le D(nv, 0) \le n\|Av\|_H + B.$$

Dividing by n yields

$$\left| \frac{D(nv,0)}{n} - \|Av\|_H \right| \le \frac{B}{n},$$

so the limit defining $\nu(v)$ exists and equals $\|Av\|_H$.

If $v \in \ker A$, then $D(nv, 0) \le B$ for all n, so $v \in E_0$. Conversely, if $Av \neq 0$, then

$$D(nv, 0) \ge n\|Av\|_H - B \to \infty,$$

so $v \notin E_0$. Hence $E_0 = \ker A$. Finally, the formula $\nu(v) = \|Av\|_H$ is constant on cosets of E_0 and is positive away from the zero coset, so it defines a norm on E/E_0.

Appendix A.10. Proof of Corollary 2

The bounded-perturbation assumption gives, for every v and $n \ge 1$,

$$\left| \frac{D'(nv,0)}{n} - \frac{D(nv,0)}{n} \right| \le \frac{B}{n}.$$

If $D(nv, 0)/n$ converges to $\nu_D(v)$, the right-hand side tends to zero, so $D'(nv, 0)/n$ has the same limit. This proves the stability of the specific energy.

If D is subadditive on triples, then

$$-\frac{D(x,z)}{\tau} \ge -\frac{D(x,y)}{\tau} - \frac{D(y,z)}{\tau}.$$

Exponentiating both sides yields

$$\exp(-D(x,z)/\tau) \ge \exp(-D(x,y)/\tau)\exp(-D(y,z)/\tau),$$

which is Equation (53).

Appendix A.11. Proof of Lemma 1

For any $t \in I$, choose a grid point $t_j \in G$ with $|t - t_j| \le \Delta/2$. Then

$$|h_v(t) - h_v(0)| \le |h_v(t) - h_v(t_j)| + |h_v(t_j) - h_v(0)| \le M\Delta/2 + \alpha,$$

which proves the claim.

Appendix A.12. Proof of Corollary 3

The quantities H, S and R are non-negative. Jensen's inequality gives

$$\mathbb{E}H \le (\mathbb{E}H^2)^{1/2} \le \sqrt{\ell_h}, \qquad \mathbb{E}S \le (\mathbb{E}S^2)^{1/2} \le \sqrt{\ell_s}, \qquad \mathbb{E}R \le (\mathbb{E}R^2)^{1/2} \le \sqrt{\ell_r}.$$

Markov's inequality applied to H^2, S^2 and R^2 yields

$$\mathbb{P}(H > t) = \mathbb{P}(H^2 > t^2) \le \frac{\mathbb{E}H^2}{t^2} \le \frac{\ell_h}{t^2},$$

and the same argument gives the bounds for S and R. Finally, $S(v, w) \le t$ is equivalent to

$$[d_\theta(v + w) - d_\theta(v) - d_\theta(w)]_+ \le t\big(d_\theta(v) + d_\theta(w) + \varepsilon\big),$$

which implies Equation (61). This proves the corollary.

Appendix A.13. Proof of Corollary 4

Conditional on the trained model and the diagnostic sampling distribution, the random variables X_i are independent, bounded in $[0, B]$ and have common mean $\mathbb{E}X$. Hoeffding's inequality gives

$$\mathbb{P}(\widehat{\mu}_M - \mathbb{E}X > \eta) \le \exp\left(-\frac{2M\eta^2}{B^2}\right), \qquad \mathbb{P}(\mathbb{E}X - \widehat{\mu}_M > \eta) \le \exp\left(-\frac{2M\eta^2}{B^2}\right).$$

Applying the union bound to the two tails yields Equation (62). This proves the corollary.

Appendix B. Hyperparameters and Experimental Details

Table A1. Core configuration of the Fashion-MNIST controlled experiment, the UCI HAR non-visual corroboration, and the auxiliary Digits check.

Item	Setting
Primary controlled image dataset	Fashion-MNIST, stratified 20,000 train/5000 test subsets
Full image split audit	Fashion-MNIST, official 60,000 train/10,000 test split, ten seeds
Post hoc image audits	Fashion-MNIST, official split, three seeds for shrinkage, residual-scale, off-grid, and latent-scale diagnostics
Secondary non-visual dataset	UCI HAR, official train/test split, standardized 561-feature vectors, ten seeds
Auxiliary image check	scikit-learn Digits, stratified splits after 28×28 preprocessing
Image encoder	Compact CNN, latent dimension 32
UCI HAR encoder	MLP, hidden dimension 64, latent dimension 32
Residual distance	$\rho_\theta(v) = \|v\|_2(1 + 0.65 \tanh g_\theta(v)) + 10^{-7}$
Fashion-MNIST AMN-balanced	$w_{\text{hom}} = 0.06$, $w_{\text{sub}} = 0.12$, $w_{\text{rec}} = 0.005$
Fashion-MNIST AMN geom-first	$w_{\text{hom}} = 0.09$, $w_{\text{sub}} = 0.08$, $w_{\text{rec}} = 0.01$
UCI HAR AMN-balanced	$w_{\text{hom}} = 0.12$, $w_{\text{sub}} = 0.24$, $w_{\text{rec}} = 0.02$
Dyadic reconstruction	$L = 4$, scales $\{1/2, 2, 4\}$ for diagnostics
Task loss	Triplet-style contrastive loss with margin $M = 1$
Reported retrieval metrics	Recall@k, nearest-neighbor accuracy and mAP where applicable

Exploratory AMN-Weight Sensitivity

The final AMN weights were fixed after an exploratory grid and before the ten-seed controlled and non-AMN-control runs. Table A2 gives a compact excerpt from the archived grid. The table is not used as a formal inference table; it documents that the selected balanced and geom-first settings lie in a stable region of the exploratory trade-off rather than being isolated single-point choices.

Table A2. Representative Fashion-MNIST AMN-weight sensitivity excerpt from the archived exploratory grid. Values are exploratory means; the final controlled run used a separate ten-seed paired run with the selected balanced and geom-first settings.

Variant	w_{hom}	w_{sub}	w_{rec}	Recall@1	D_{hom}	D_{sub}	D_{rec}
low hom/sub	0.03	0.08	0.000	0.8762	0.196	0.0102	0.195
mid hom/sub	0.06	0.08	0.000	0.8775	0.164	0.0084	0.154
mid + rec	0.06	0.08	0.005	0.8764	0.159	0.0082	0.148
balanced selected	0.06	0.12	0.005	0.8780	0.159	0.0082	0.149
geom-first selected	0.09	0.08	0.010	0.8757	0.132	0.0063	0.118
strong no-rec	0.09	0.12	0.000	0.8750	0.139	0.0068	0.127

The residual family was selected because fully flexible heads were unstable in exploratory phases. In the Fashion-MNIST controlled experiment, the same overall training protocol was held fixed, and only the AMN-inspired or control weights were changed across variants. The additional Fashion-MNIST theory audits reuse the official full split for three seeds to test Euclidean shrinkage, the residual-scale parameter a, off-training scales, and latent-operation extrapolation; they are not used for hyperparameter selection. The UCI HAR corroboration uses the same residual function and diagnostic definitions with a compact MLP encoder on standardized 561-feature vectors; it is reported as a secondary ten-seed CPU run with full-query/full-gallery evaluation, not as a separate hyperparameter search. The Digits experiment reuses the convolutional code path after deterministic preprocessing to a 28×28 image format and is retained as an auxiliary image-domain check.

All final results are computed from seed-level CSV files included in the supplementary package. The supplementary analysis files also contain paired t-test and Wilcoxon signed-rank summaries against the residual task-only baseline. These tests are used descriptively to check seed-level consistency; the paper does not rely on a binary significance threshold.

Appendix C. Additional Supporting Results

Table A3. Supporting MNIST/Fashion-MNIST results from the pre-final phase. These runs helped stabilize the mechanism before the final Fashion-MNIST controlled experiment and the non-AMN controls.

Dataset	Variant	Recall@1	D_{hom}	D_{sub}	D_{rec}
MNIST	Residual task-only	0.9916	0.0860	0.0018	0.1266
MNIST	Residual hom + sub	0.9912	0.0746	0.0040	0.0699
Fashion-MNIST	Residual task-only	0.8850	0.2114	0.0086	0.2342
Fashion-MNIST	Residual hom + sub	0.8823	0.1216	0.0065	0.1310

The auxiliary scikit-learn Digits benchmark and the robustness probes are kept in this appendix so that the main text remains focused on the primary Fashion-MNIST and UCI HAR evidence.

Table A4. Auxiliary scikit-learn Digits pilot. Values are archived pilot means and standard deviations over five seeds. The table is interpreted as an additional mechanism check, not as a full large-scale validation.

Variant	Recall@1	D_{hom}	D_{sub}	D_{rec}
Euclidean triplet	0.9796 ± 0.0056	0.0007 ± 0.0003	0.000000 ± 0.000000	0.0006 ± 0.0002
Residual task-only	0.9780 ± 0.0070	0.0389 ± 0.0049	0.000247 ± 0.000037	0.0329 ± 0.0062
Homogeneity + subadditivity	0.9791 ± 0.0055	0.0090 ± 0.0018	0.000706 ± 0.000634	0.0152 ± 0.0056
AMN reconstruction	0.9746 ± 0.0058	0.0141 ± 0.0188	0.000033 ± 0.000059	0.0310 ± 0.0449

The residual task-only model already has high Recall@1 on this easier handwritten-digit task. The homogeneity-plus-subadditivity variant keeps the retrieval score essentially unchanged while reducing D_{hom} and D_{rec}. The absolute subadditivity defect is extremely small in the task-only baseline; percentage changes in D_{sub} should therefore not be interpreted.

Table A5. Secondary robustness probes. Entries are Recall@1 under the listed perturbation. The table is diagnostic and is not a central contribution.

Variant	Gaussian	Occlusion	Contrast	Shift
Residual task-only	0.2551	0.5199	0.8371	0.6542
AMN-balanced	0.2962	0.5025	0.8327	0.6575
AMN geom-first	0.2947	0.5016	0.8326	0.6543

Robustness results are mixed: AMN-inspired variants improve the Gaussian-noise probe relative to residual task-only, are close on shift, and are slightly worse on occlusion and contrast. They are therefore retained only as secondary diagnostics.

References

1. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Washington, DC, USA, 2006; Volume 2, pp. 1735–1742.
2. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised Contrastive Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18661–18673.
3. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Washington, DC, USA, 2015; pp. 815–823.
4. Bellet, A.; Habrard, A.; Sebban, M. A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv* **2013**, arXiv:1306.6709.
5. Musgrave, K.; Belongie, S.; Lim, S.-N. A Metric Learning Reality Check. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 681–699.
6. Muñoz-García, E. Rigidity of AMN Vector Spaces. *arXiv* **2000**, arXiv:math/0008095.
7. Berg, C.; Christensen, J.P.R.; Ressel, P. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*; Springer: New York, NY, USA, 1984.
8. Schölkopf, B.; Smola, A.J. *Learning with Kernels*; MIT Press: Cambridge, MA, USA, 2002.
9. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
10. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
11. Weinberger, K.Q.; Saul, L.K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *J. Mach. Learn. Res.* **2009**, *10*, 207–244.
12. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Washington, DC, USA, 2005; pp. 539–546.
13. Kim, S.; Kim, D.; Cho, M.; Kwak, S. Proxy Anchor Loss for Deep Metric Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Washington, DC, USA, 2020; pp. 3238–3247. [CrossRef]

14. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle Loss: A Unified Perspective of Pair Similarity Optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Washington, DC, USA, 2020; pp. 6398–6407.
15. Mohan, D.D.; Jawade, B.; Setlur, S.; Govindaraju, V. Deep Metric Learning for Computer Vision: A Brief Overview. In *Deep Learning*; Handbook of Statistics; Elsevier: Amsterdam, The Netherlands, 2023; Volume 48, pp. 59–79.
16. Kaya, M.; Bilge, H.Ş. Deep Metric Learning: A Survey. *Symmetry* **2019**, *11*, 1066. [CrossRef]
17. Li, X.; Yang, X.; Ma, Z.; Xue, J.-H. Deep Metric Learning for Few-Shot Image Classification: A Review of Recent Developments. *Pattern Recognit.* **2023**, *138*, 109381. [CrossRef]
18. Hoeffding, W. Probability Inequalities for Sums of Bounded Random Variables. *J. Am. Stat. Assoc.* **1963**, *58*, 13–30. [CrossRef]
19. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv* **2017**, arXiv:1708.07747.
20. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
21. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
22. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition Using Smartphones. In *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*; ESANN: Bruges, Belgium, 2013; pp. 437–442.
23. Reyes-Ortiz, J.L.; Anguita, D.; Ghio, A.; Oneto, L.; Parra, X. Human Activity Recognition Using Smartphones. In *UCI Machine Learning Repository*; University of California, Irvine: Irvine, CA, USA, 2013. [CrossRef]
24. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.