

HyDRA: Hyperbolic Damage-aware Recovery and Adaptation for Robust Multimodal Sentiment Analysis with Incomplete Data

Yun Luo^{1,2,3†}, Zejiang He^{1,2†}, Zhen Huang^{1,2*}, Zhiliang Tian^{1,2*},
Yanzhou Ma³, Menglong Lu^{1,2}, Jingyuan Huang^{1,2},
Shanshan Liu^{1,2}, Faying Wu⁴, Dongsheng Li^{1,2}

¹College of Computer Science and Technology, National University of Defense Technology,

²National Key Laboratory of Parallel and Distributed Computing,

³College of Foreign Languages, National University of Defense Technology,

⁴School of Information and Electronics, Beijing Institute of Technology

{hezejiang, ly24, huangzhen, tianzhiliang, jingyuanhuang, liushanshan17, myz827, dsli}@nudt.edu.cn

lumenglong1995@163.com, 3220250899@bit.edu.cn

Abstract

Incomplete-Modality Multimodal Sentiment Analysis extends multimodal sentiment analysis (MSA) to infer sentiment from language, audio, and visual streams when one or more modalities are incomplete at inference time. Existing methods improve incomplete-modality MSA through feature reconstruction, distillation, or fusion adjustment, but usually assign similar correction strength to all samples with incomplete inputs. This overlooks the uneven emotional impact of modality missingness: mild missingness often leaves sentiment recognizable, whereas severe missingness can distort the expressed emotion and therefore requires stronger correction. We propose **HyDRA** (**H**yperbolic **D**amage-aware **R**ecovery and **A**daptation), a damage-aware and text-guided hyperbolic framework that aligns correction strength with missingness-induced sentiment damage. HyDRA first recovers lost emotional cues from available modalities, then places each sample in the Poincaré ball according to estimated damage severity, and finally applies damage-weighted hyperbolic correction toward a complete-modality teacher anchor. Since hyperbolic distances expand near the boundary, heavily incomplete samples receive stronger correction while lightly incomplete samples are gently adjusted. Experiments on MOSI, MOSEI, and SIMS show SOTA robustness across incomplete-modality conditions.

1 Introduction

Incomplete-Modality Multimodal Sentiment Analysis extends MSA to infer sentiment from language,

[†] contributed equally to this work

* corresponding authors

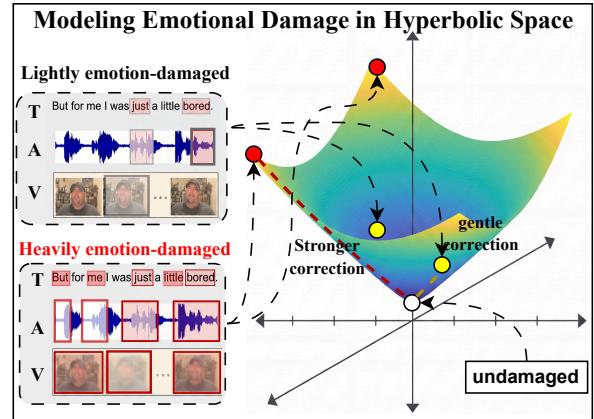


Figure 1: Samples with complete modalities reside near the origin, while those affected by modality missingness are pushed toward the boundary. In the Poincaré ball, distances are naturally amplified closer to the boundary, so heavily damaged samples receive stronger correction, while lightly damaged samples are gently adjusted.

audio, and visual streams when one or more modalities are partially or entirely unavailable. This setting is common in real deployments, where sensor failures, network interruptions, or automatic speech recognition (ASR) errors can produce incomplete inputs (Yuan et al., 2021; Zhang et al., 2024a). Although complete-input MSA has made strong progress with cross-modal interaction strategies (Zadeh et al., 2017; Liang et al., 2021; Han et al., 2021; Wang et al., 2022, 2023; Zhang et al., 2023), its complete-modality assumption limits robustness under incomplete inputs.

Existing research on this task can be broadly categorized into three lines: (1) *Feature reconstruction methods* (Yuan et al., 2021; Yang et al., 2025; Zhu et al., 2025) recover unavailable modality fea-

tures, using cross-modal generative mappings or learned proxy features; (2) *Distillation-based methods* (Ma et al., 2021; Li et al., 2024; Luo et al., 2025) leverage a complete-modality teacher network to supervise a student network that handles incomplete inputs, thereby avoiding explicit feature imputation; (3) *Fusion-adjustment methods* (Zhang et al., 2024a) re-weight each modality’s contribution during fusion by assigning completeness scores.

Although these methods tackle modality missingness from different angles, they all overlook a key fact: **the impact of modality missingness varies sharply across samples and missing rates**. Intuitively, losing only a small part of a modality usually leaves a sample’s emotion recognizable, whereas losing most of it can seriously distort what the sample expresses. This suggests that a model should apply stronger corrective guidance to more severely damaged samples rather than using a correction strength that stays fixed (Nickel and Kiela, 2017; Ganea et al., 2018; Chami et al., 2019). Our empirical analysis (Appendix A.1) confirms this intuition: lightly incomplete samples remain close to their complete-input representations, whereas heavily incomplete samples move much farther away and are harder to correct with a single correction strength. Existing methods, however, cannot accommodate such dynamics. As shown in Figure 1, hyperbolic geometry offers a principled solution: in the Poincaré ball, negative curvature causes distances to expand substantially near the boundary, naturally imposing **stronger correction on heavily damaged samples** while **gently adjusting lightly damaged samples**.

In this paper, we propose **HyDRA** (Hyperbolic Damage-aware Recovery and Adaptation), a damage-aware and text-guided hyperbolic framework for Incomplete-Modality Multimodal Sentiment Analysis that matches correction strength to the emotional damage caused by modality missingness. Specifically, to recover sentiment cues lost before fusion, we propose **Cross-Modal Emotional Recovery (CMER)** that retrieves audio-visual evidence with learnable proxy tokens and injects it into the language stream through adaptive gating. To further organize samples by damage severity, we introduce **Damage-Aware Hyperbolic Positioning (DAHP)**, which converts per-sample completeness into a radial coordinate in the Poincaré ball. DAHP decouples sentiment semantics from damage severity by using the direction to encode

text-guided fused content and the radius to encode reliability, placing low-completeness samples closer to the boundary where hyperbolic distances are naturally amplified. To adapt correction strength during training, we design **Severity-Adaptive Hyperbolic Correction (SAHC)** that transfers complete-modality teacher knowledge with a damage-weighted hyperbolic distance, so low-completeness samples placed near the boundary receive stronger correction while reliable samples remain stable. Experiments on three benchmark datasets under missing rates from 0 to 0.9 show that HyDRA achieves state-of-the-art robustness, with up to a 48.7% relative Acc-7 gain over P-RMF under severe missingness.

The main contributions of this paper are: (1) We propose HyDRA, a damage-aware hyperbolic framework for Incomplete-Modality Multimodal Sentiment Analysis. We model modality missingness as sample-level continuous damage and use completeness-controlled radial positioning to distinguish reliable representations from severely damaged ones in the Poincaré ball. (2) We propose a severity-adaptive recovery and correction mechanism by coupling cross-modal emotional recovery with boundary-amplified hyperbolic distillation. This design restores missing sentiment cues before fusion and assigns stronger teacher-guided correction to lower-completeness samples. (3) Experiments on MOSI, MOSEI, and SIMS demonstrate state-of-the-art robustness across various incomplete-modality conditions.

2 Related Work

2.1 Multimodal Sentiment Analysis

MSA methods usually assume complete language, audio, and visual streams and focus on stronger cross-modal representations. Representative work models high-order fusion, sequence alignment, modality-invariant/specific factors, and non-verbal augmentation for pretrained language models (Zadeh et al., 2017; Tsai et al., 2019; Hazarika et al., 2020; Rahman et al., 2020). Subsequent studies improve supervision, alignment, and fusion through self-supervision, mutual information, cross-modal enhancement, text-enhanced Transformers, language-guided hyper-modality representations, asynchronous regularization, unified sentiment/emotion modeling, emotion-aware fusion, and unimodal label correction (Yu et al., 2021; Han et al., 2021; Wang et al., 2022, 2023; Zhang et al.,

2023; Liang et al., 2021; Hu et al., 2022; Yuan et al., 2024; Zhang et al., 2024b). These works provide strong complete-input backbones, but they do not specify how to recover or correct samples when arbitrary portions of a modality are unavailable at inference time for a test sample.

2.2 Incomplete-Modality Multimodal Sentiment Analysis

Incomplete-modality MSA studies sentiment prediction when modality streams are partially missing or corrupted. Under the common intra-modal protocol, a missing rate r_m controls how much of each stream is erased. Existing methods improve robustness through feature recovery, modality weighting, or teacher guidance: TFR-Net reconstructs missing features under varying missing rates (Yuan et al., 2021); SMIL addresses missing modality combinations under a different protocol (Ma et al., 2021); LNLN reweights modalities during fusion (Zhang et al., 2024a); P-RMF uses proxy-driven recovery for incomplete language (Zhu et al., 2025); and DAR decouples sentiment-related and irrelevant factors before reconstruction (Yang et al., 2025). Recent work further explores confidence-aware self-distillation, foundation-model-assisted correction, contrastive multimodal interaction graphs, and meta-learned modality weighting (Luo et al., 2025; Zhao et al., 2025; Gong et al., 2026; Tu et al., 2025). Although these methods form strong baselines, their recovery, distillation, or fusion signals are usually not explicitly controlled by each sample’s estimated damage severity. HyDRA instead maps each sample to a damage-controlled geometric position and uses that position to modulate correction strength.

2.3 Hyperbolic Geometry in Representation Learning

Hyperbolic geometry is commonly used to encode non-uniform structure because distances expand rapidly near the boundary, a property that is attractive when samples have different damage levels. Poincaré and Lorentz embeddings, hyperbolic neural networks, attention, graph neural networks, and image embeddings show that curved spaces can represent hierarchy or structured similarity in compact dimensions (Nickel and Kiela, 2017, 2018; Ganea et al., 2018; Gulcehre et al., 2019; Chami et al., 2019; Liu et al., 2019; Khruikov et al., 2020; Desai and Johnson, 2021). HyDRA also draws on knowledge distillation, where a teacher-

student framework transfers information from a stronger model to a compact or incomplete student (Hinton et al., 2015). Existing hyperbolic models mainly use curvature for representation, while distillation-based incomplete-modality methods usually lack a sample-level geometric control signal. HyDRA connects these lines by placing samples at completeness-controlled radii and turning the Poincaré ball’s boundary amplification into adaptive teacher correction.

3 Method

3.1 Overall Framework

Given multimodal inputs containing language, audio, and visual modalities, Incomplete-Modality Multimodal Sentiment Analysis aims to predict the sentiment label $y \in \mathbb{R}$ for regression or $y \in \{1, \dots, C\}$ for classification when arbitrary portions of each modality are unavailable. For each modality $m \in \{l, a, v\}$ (language, audio, and visual), let $U^m \in \mathbb{R}^{T_m \times d_m}$ denote the complete feature sequence and $\tilde{U}^m$ denote its incomplete counterpart after masking. The missingness simulation protocol is detailed in § 4.1.

To address missingness-induced damage that grows with the amount of lost sentiment evidence, HyDRA takes incomplete multimodal sequences as input and predicts sentiment by recovering missing cues, positioning each sample in a damage-aware hyperbolic space, and applying severity-adaptive teacher correction. As shown in Figure 2, the data flow is $\tilde{U}^m \rightarrow H^m \rightarrow \hat{H}^l \rightarrow (\hat{w}, G, \hat{z}, z) \rightarrow \mathcal{L}_{\text{hyp}}$. The three modules of HyDRA are designed to make correction strength sample-adaptive rather than uniform. **Cross-Modal Emotional Recovery (CMER)** (§ 3.2) first reconstructs missing sentiment evidence at the representation level by using learnable proxy tokens to query audio-visual streams and gatedly inject the retrieved evidence into language features. **Damage-Aware Hyperbolic Positioning (DAHP)** (§ 3.3) explicitly assigns each sample a damage-aware location in the Poincaré ball with a disentangled geometry: the angular direction carries text-guided sentiment semantics, whereas the radial position is controlled by completeness and therefore encodes damage severity. **Severity-Adaptive Hyperbolic Correction (SAHC)** (§ 3.4) further exploits the boundary expansion of the Poincaré ball by applying damage-weighted teacher-student correction, making severely incomplete samples receive amplified

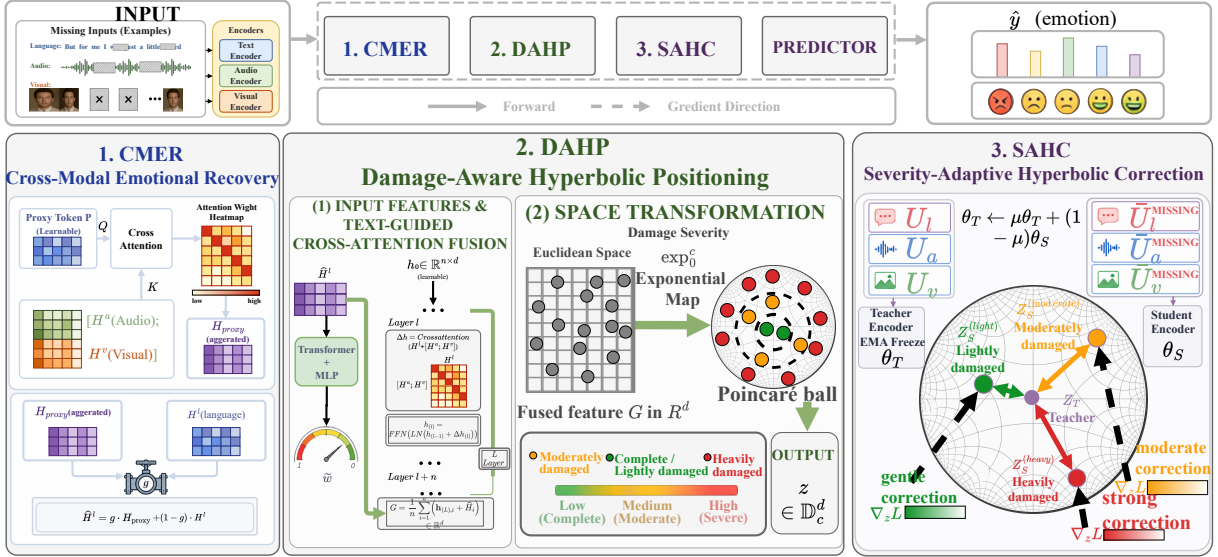


Figure 2: Overall framework of HyDRA: **Cross-Modal Emotional Recovery** restores the emotional cues lost to modality missingness, **Damage-Aware Hyperbolic Positioning** aggregates the recovered features and projects them onto the Poincaré ball so that each sample’s position reflects its damage severity, and **Severity-Adaptive Hyperbolic Correction** transfers teacher knowledge with correction strength matched to each sample’s damage severity.

guidance and keeping mildly incomplete samples stable.

3.2 Cross-Modal Emotional Recovery

To recover missing emotional cues before fusion, we propose **Cross-Modal Emotional Recovery (CMER)** to compensate the language stream on incomplete-modality inputs, which retrieves audio-visual evidence with learnable proxy tokens and injects it through adaptive gating. Since language usually carries the main sentiment semantics in MSA (Zhang et al., 2023, 2024a), directly replacing unavailable language features with zeros or fixed values can severely degrade the fused representation (Yuan et al., 2021; Zhu et al., 2025).

Specifically, we first encode each incomplete modality $m \in \{l, a, v\}$ with a modality-specific encoder $\mathcal{E}^m$:

$$H^m = \mathcal{E}^m(\tilde{U}^m) \in \mathbb{R}^{n \times d} \quad (1)$$

where n is the unified token length and d is the hidden dimension, so the three streams can be processed in a shared feature space. Then, to recover language cues that are lost in $\tilde{U}^l$, a learnable proxy token $\mathbf{P} \in \mathbb{R}^{n \times d}$ is used as the query over the concatenated audio-visual features $[H^a; H^v]$, yielding a language-oriented proxy representation H_{proxy} . Finally, to decide how much proxy information should be injected, we adaptively fuse H_{proxy} with

the encoded language features H^l through scalar gating:

$$\hat{H}^l = g \cdot H_{\text{proxy}} + (1 - g) \cdot H^l \quad (2)$$

where $g = \sigma(W_g[\bar{H}_{\text{proxy}}; \bar{H}^a; \bar{H}^v] + b_g)$ is computed from the sequence-averaged features with sigmoid activation, and $\bar{H} = \frac{1}{n} \sum_{i=1}^n H_i$ denotes the sequence-averaged representation of an arbitrary feature matrix $H \in \mathbb{R}^{n \times d}$. We use gated proxy injection because the gate g adaptively controls the amount of injected proxy evidence: when the language stream is heavily corrupted, the model relies more on the audio-visual proxy H_{proxy} ; when the language stream remains reliable, it preserves more of the original representation H^l . The compensated representation $\hat{H}^l$ therefore provides the semantic input for DAHP.

3.3 Damage-Aware Hyperbolic Positioning

To connect correction strength with damage severity, we propose **Damage-Aware Hyperbolic Positioning (DAHP)** to place each fused representation in the Poincaré ball, where the direction is computed from sentiment-relevant fused features, and the distance from the origin is controlled by estimated completeness; thus, lower-completeness samples are placed farther from the center and treated as needing stronger correction. A key limitation of existing fusion strategies is that the fused

representation does not explicitly indicate whether a sample is mildly or severely incomplete (Zhu et al., 2025; Yang et al., 2025). DAHP addresses this by encoding reliability as the radial position in the Poincaré ball, so the subsequent correction can distinguish samples that require weak or strong guidance. Specifically, DAHP first estimates a completeness score $\hat{w}$, then builds a semantic proposal $\tilde{z}$ from recovered language evidence, and finally adjusts only its radius to obtain z in a severity-aware Poincaré ball; formal definitions are in Appendix H.

(1) Missingness-Aware Completeness Scoring.

The completeness score is designed to tell the model how reliable the current sample is before it enters hyperbolic correction. Since language usually carries the dominant sentiment semantics in MSA, the language missing rate provides a direct indicator of how much sentiment evidence has been lost. We therefore summarize the recovered language features $\hat{H}^l$ with a shallow Transformer $\mathcal{T}_{\text{comp}}$ and use its CLS token $\hat{h}_{\text{cls}}^l$ as the sample state. A two-layer MLP maps this state to $\hat{w} \in [0, 1]$, supervised by $w = 1 - r_l$ following Zhang et al. (2024a), where r_l is the language missing rate in Equation (19). In this way, $\hat{w}$ acts as a missingness-aware reliability signal: high $\hat{w}$ indicates that the language evidence is mostly preserved, while low $\hat{w}$ indicates severe semantic damage and triggers larger radial displacement and stronger correction in the following modules.

(2) Text-Guided Semantic Direction Construction. After estimating sample reliability, DAHP needs to determine where the sample should point in the Poincaré ball. We use the recovered language representation as the semantic anchor because language carries the main sentiment content, and retrieve complementary audio-visual cues through cross-attention. This produces a text-guided semantic representation G , which is then mapped to the Poincaré ball as the proposal $\tilde{z}$. The direction of $\tilde{z}$ represents sentiment semantics, while its radius will be adjusted by the completeness score in the next step. Specifically, we initialize a learnable semantic guidance token $\mathbf{h}_{(0)} \in \mathbb{R}^{n \times d}$; at each layer ℓ , $\mathbf{h}_{(\ell-1)}$ queries the audio and visual features, and the update $\Delta \mathbf{h}_{(\ell)}$ is refined as

$$\mathbf{h}_{(\ell)} = \text{FFN}(\text{LN}(\mathbf{h}_{(\ell-1)} + \Delta \mathbf{h}_{(\ell)})) \quad (3)$$

where $\Delta \mathbf{h}_{(\ell)}$ denotes the sum of cross-attention outputs from auxiliary modalities $\{a, v\}$. We average

the final guidance features with $\hat{H}^l$ to obtain

$$G = \frac{1}{n} \sum_{i=1}^n (\mathbf{h}_{(L),i} + \hat{H}_i^l) \in \mathbb{R}^d \quad (4)$$

and map G onto the Poincaré ball as the semantic proposal $\tilde{z}$:

$$\tilde{z} = \exp_0^c(W_{\text{hyp}} G + b_{\text{hyp}}) \in \mathbb{D}_c^{d'}. \quad (5)$$

where $W_{\text{hyp}} \in \mathbb{R}^{d' \times d}$ and $b_{\text{hyp}} \in \mathbb{R}^{d'}$ are learnable parameters.

(3) Completeness-Controlled Radial Positioning. Given the semantic proposal $\tilde{z}$, DAHP keeps its direction fixed and changes only its radius according to the completeness score $\hat{w}$. This realizes damage-aware positioning: high-completeness samples stay near their original semantic proposal, while low-completeness samples are pushed closer to the Poincaré boundary, where subsequent hyperbolic correction is amplified.

$$u = \frac{\tilde{z}}{\|\tilde{z}\|_2 + \epsilon}, \quad (6)$$

$$\rho_{\max} = \frac{1 - \epsilon}{\sqrt{c}}, \quad (7)$$

$$\rho(\hat{w}) = \hat{w} \|\tilde{z}\|_2 + (1 - \hat{w}) \rho_{\max}, \quad (8)$$

$$z = \Pi_c(\rho(\hat{w}) u) \in \mathbb{D}_c^{d'}. \quad (9)$$

where ϵ is used for numerical stability and Π_c is boundary projection. We keep the direction u fixed and change only the radius because the direction carries sentiment content, while the radius should reflect how strongly the sample needs correction; a symbol table and an intuitive reading are in Appendix H.8.

3.4 Severity-Adaptive Hyperbolic Correction

To adapt correction strength to damage severity, we propose **Severity-Adaptive Hyperbolic Correction (SAHC)** to transfer complete-modality teacher knowledge on the Poincaré ball, which weights the teacher-student hyperbolic discrepancy by estimated incompleteness after DAHP positioning. Uniform distillation assigns similar correction strength to all samples with incomplete inputs (Li et al., 2024; Luo et al., 2025), which can under-correct heavily damaged samples and disturb lightly damaged samples.

Specifically, we first maintain a complete-modality teacher as an exponential moving average (EMA) of the student parameters:

$$\theta_T \leftarrow \mu \theta_T + (1 - \mu) \theta_S \quad (10)$$

where θ_S and θ_T are the parameters of the student and teacher models, respectively, and μ is the momentum coefficient. The teacher receives complete multimodal inputs to produce the hyperbolic anchor z_T , while the student receives incomplete inputs to produce z_S . Since z_T is produced from complete inputs, it serves as the target anchor that the incomplete representation z_S should approach. We therefore measure their discrepancy with the Poincaré distance. This is not only a distance metric: because low-completeness samples are placed near the Poincaré boundary by DAHP, their deviation from the teacher is geometrically amplified, thereby aligning the correction strength with the estimated severity of sentiment-evidence loss. We then weight this discrepancy by $1 - \hat{w}$:

$$\mathcal{L}_{\text{hyp}} = \max(0, d_{\mathbb{H}}(z_S, z_T) - \tau) \cdot (1 - \hat{w}) \quad (11)$$

where τ is a relaxation margin and $d_{\mathbb{H}}$ is the hyperbolic distance (see Appendix H). The hinge term ignores pairs already close to the teacher, and the factor $1 - \hat{w}$ makes low-completeness samples receive stronger teacher guidance.

We weight the teacher–student hyperbolic discrepancy by $1 - \hat{w}$ because this directly ties correction strength to estimated damage severity: low-completeness samples receive stronger teacher guidance, while high-completeness samples are less disturbed.

3.5 Overall Learning Objective

To train HyDRA end to end, we use an overall learning objective that jointly optimizes sentiment prediction and damage-aware correction for Incomplete-Modality MSA. The objective consists of four terms: a task loss, a completeness loss, a hyperbolic distillation loss, and a reconstruction loss. Specifically, the task loss $\mathcal{L}_{\text{task}}$ trains the final sentiment predictor. The completeness loss $\mathcal{L}_{\text{comp}} = \|\hat{w} - w\|^2$ supervises the completeness score using the target w defined in § 3.3. The hyperbolic distillation loss $\mathcal{L}_{\text{hyp}}$ in Equation (11) provides severity-adaptive teacher guidance. To keep modality-level information available during representation learning, the reconstruction loss $\mathcal{L}_{\text{recon}}$ uses lightweight decoders $\mathcal{D}^m$:

$$\mathcal{L}_{\text{recon}} = \sum_{m \in \{l, a, v\}} \|\mathcal{D}^m(H^m) - U^m\|_2^2 \quad (12)$$

Finally, these terms are combined as:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_1 \mathcal{L}_{\text{comp}} + \beta(t) \mathcal{L}_{\text{hyp}} + \lambda_2 \mathcal{L}_{\text{recon}} \quad (13)$$

where λ_1 and λ_2 are balancing hyperparameters and $\beta(t)$ is a progressive distillation weight. Considering the sensitivity of hyperbolic optimization in early stages, $\beta(t)$ remains zero for the first T_{warmup} training epochs and then increases linearly to β_{max} . This schedule lets the model first learn stable feature representations and then introduces teacher correction after the Poincaré ball representation becomes stable.

4 Experiments

4.1 Experiment Setups

Datasets. We evaluate on CMU-MOSI (Zadeh et al., 2016), CMU-MOSEI (Zadeh et al., 2018), and SIMS (Yu et al., 2020); dataset statistics and descriptions are provided in Appendix G.1.

Baselines. We compare with (1) *complete-modality methods*: MISA (Hazariika et al., 2020), Self-MM (Yu et al., 2021), MMIM (Han et al., 2021), CENET (Wang et al., 2022), TETFN (Wang et al., 2023), and ALMT (Zhang et al., 2023); (2) *incomplete-modality methods*: TFR-Net (Yuan et al., 2021), LNLN (Zhang et al., 2024a), and P-RMF (Zhu et al., 2025). Detailed baseline descriptions are provided in Appendix G.2.

Evaluation and Missing Protocol. We follow Zhang et al. (2024a) and evaluate missing rates $r \in \{0.0, 0.1, \dots, 0.9\}$. The exact mask construction, modality-specific fill values, and formal missing-input definition are given in Appendix G.3. For MOSI and MOSEI, Acc-2/F1 follow the standard Has0/Non0 convention, detailed in Appendix G.3.

Implementations. Feature extraction, optimizer settings, full hyperparameters, and the independent five-seed significance protocol are reported in Appendices G.4 and G.5; sensitivity analyses are in Appendix F.

4.2 Overall Performance

Table 1 reports the overall performance of HyDRA and the compared methods on MOSI, MOSEI, and SIMS under incomplete-modality conditions. The results show that HyDRA surpasses all baselines on most metrics across the three datasets, demonstrating strong overall robustness to modality missingness. Noticeably, HyDRA achieves the largest improvement on MOSI Acc-7, with a 7.6% relative gain over the second-best baseline, indicating its advantage in fine-grained sentiment prediction.

Model	MOSI						MOSEI						SIMS					
	Acc-2	F1	Acc-5	Acc-7	MAE↓	Corr↑	Acc-2	F1	Acc-5	Acc-7	MAE↓	Corr↑	Acc-5	Acc-3	Acc-2	F1	MAE↓	Corr↑
MISA	71.5/70.3	71.3/70.0	33.08	29.85	1.085	0.524	71.3/75.8	63.9/68.7	39.39	40.84	0.780	0.503	31.53	56.87	72.71	66.30	0.539	0.348
Self-MM	70.5/69.3	66.6/67.5	34.68	29.55	1.070	0.512	73.9/77.4	68.9/72.3	45.38	44.70	0.695	0.499	32.28	56.75	72.81	68.43	0.508	0.376
MMIM	69.1/67.1	66.7/64.0	33.77	31.30	1.077	0.507	73.3/75.9	68.7/70.3	41.73	40.75	0.739	0.489	31.81	52.76	69.86	66.21	0.544	0.339
CENET	71.5/67.7	68.4/64.9	37.25	30.38	1.080	0.504	74.7/77.3	70.7/74.1	47.83	47.18	0.685	0.535	22.29	53.17	68.13	57.90	0.589	0.107
TETFN	69.8/67.7	65.7/63.3	34.34	30.30	1.087	0.507	73.5/76.8	68.9/72.5	45.92	44.18	0.702	0.521	33.42	56.91	73.58	68.67	0.505	0.387
TFR-Net	68.2/66.4	61.7/60.1	34.67	29.54	1.200	0.459	73.6/77.2	68.8/72.0	34.67	46.82	0.697	0.489	26.52	52.89	68.13	58.70	0.661	0.169
ALMT	68.4/70.4	64.3/67.9	33.25	29.59	1.089	0.513	77.1/75.6	74.3/72.1	44.82	43.97	0.716	0.543	34.32	56.41	71.45	69.94	0.523	0.365
LNLN	71.4/71.2	71.3/71.1	36.42	<u>32.29</u>	1.069	0.499	76.4/77.5	75.8/77.4	46.55	45.70	0.684	0.554	32.31	56.88	72.87	69.32	0.530	0.359
P-RMF	71.4/70.4	71.2/70.3	<u>38.82</u>	31.26	1.085	0.512	<u>78.4/78.9</u>	<u>78.0/77.8</u>	<u>48.61</u>	<u>47.49</u>	<u>0.658</u>	<u>0.596</u>	<u>34.91</u>	56.56	72.89	<u>71.30</u>	0.509	<u>0.405</u>
HyDRA	71.8/73.2	71.6/73.2	38.90	34.73	1.044	0.539	78.9/78.5	78.1/78.1	49.08	48.14	0.656	0.600	35.51	58.51	<u>73.48</u>	71.96	0.496	0.430

Table 1: Robustness comparison under intra-modal missingness on MOSI, MOSEI, and SIMS. The best and second-best results are in **bold** and underlined. For MOSI and MOSEI, each Acc-2/F1 cell is reported as Has0/Non0. Our improvements are statistically significant under the t -test with $p < 0.05$ (see details in Appendix E.1).

On MOSEI Acc-2 and SIMS Acc-2, HyDRA is slightly lower than the best baseline, one reason is that these two settings rely more on dominant polarity cues while HyDRA focuses on recovering missingness-induced sentiment damage.

Figure 3 further compares the degradation trends of HyDRA and representative baselines as the missing rate increases from 0 to 0.9 on MOSI, MOSEI, and SIMS. Across the three datasets, HyDRA generally maintains higher accuracy and lower error, with larger performance gap at medium and high missing rates, suggesting ours stronger robustness when missing modalities cause larger sentiment damage. Detailed per-rate results and efficiency analysis are provided in Appendices D and I.

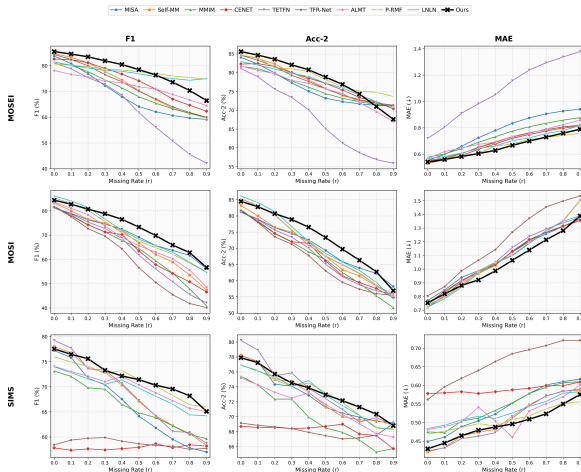


Figure 3: Performance trends of different models under missing rates 0–0.9 on MOSI, MOSEI, and SIMS. HyDRA maintains superior robustness, particularly at high missing rates.

4.3 Ablation Study

To verify the contribution of each component, we conduct ablation study on MOSI and SIMS: **w/o**

CMER removes Cross-Modal Emotional Recovery (§ 3.2); **w/o DAHP** removes Damage-Aware Hyperbolic Positioning (§ 3.3); **w/o Certainty** removes the completeness signal used in DAHP (§ 3.3); **w/o Hyp. Metric** removes the hyperbolic metric in SAHC (§ 3.4); and **w/o SAHC** removes Severity-Adaptive Hyperbolic Correction (§ 3.4).

As shown in Table 2, each component is essential to HyDRA. Removing CMER weakens cross-modal compensation, while removing DAHP or the completeness signal reduces the model’s ability to distinguish different missingness severities. The drops from **w/o Hyp. Metric** and **w/o SAHC** further show that hyperbolic teacher correction provides useful severity-adaptive guidance. The only exception is SIMS Acc-5, where **w/o CMER** is slightly higher, suggesting that this fine-grained five-level metric may be more sensitive to local decision-boundary calibration. Additional controlled analyses are provided in Appendix C.1, and qualitative examples are given in Appendix J.

4.4 Representation Visualization

To validate HyDRA’s representation recovery, we conduct a UMAP visualization experiment to compare original language features with corrected features under different missing rates. As shown in Figure 4, corrected features largely overlap with the original $r = 0$ features under moderate missingness, indicating effective alignment with the complete-data manifold. Separation appears mainly under extreme textual missingness, where limited language evidence remains. Overall, HyDRA keeps recovered representations close to the complete-data manifold across missing-rate settings, preserving semantic consistency rather than merely fitting incomplete inputs. Detailed silhouette scores of UMAP are provided in Appendix B.1

Variant	MOSI						SIMS					
	Acc-7	Acc-5	Acc-2	F1	MAE↓	Corr↑	Acc-5	Acc-3	Acc-2	F1	MAE↓	Corr↑
Ours	34.73	38.90	71.80/73.21	71.61/73.16	1.044	0.539	35.51	58.51	73.48	71.96	0.496	0.430
w/o CMER	34.60	38.68	71.58/72.96	71.42/72.88	1.059	0.535	35.62	57.95	73.02	71.88	0.515	0.415
w/o DAHP	34.22	38.30	71.15/72.52	71.02/72.42	1.078	0.518	34.85	57.62	72.68	71.18	0.518	0.408
w/o Certainty	34.28	38.22	71.05/72.48	70.92/72.35	1.088	0.515	34.72	57.52	72.55	71.05	0.525	0.402
w/o Hyp. Metric	34.05	38.08	71.12/72.45	70.95/72.32	1.092	0.510	34.58	57.28	72.42	70.82	0.532	0.398
w/o SAHC	33.92	37.85	70.75/72.15	70.65/72.05	1.128	0.512	34.95	57.73	72.96	71.45	0.522	0.408

Table 2: Component ablation study on MOSI and SIMS. Each **w/o** variant removes the corresponding HyDRA component. For MOSI, each Acc-2/F1 cell is reported as Has0/Non0. **Bold**: best; underline: second best.

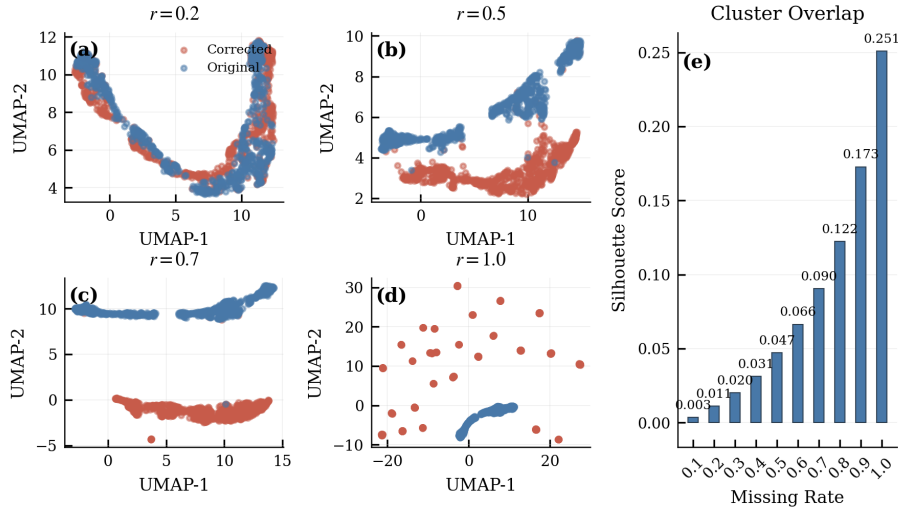


Figure 4: UMAP projections of original language features (blue, $r = 0$) and corrected features (coral) under increasing missing rates (a)–(d), alongside silhouette scores (e) quantifying cluster separability. Scores near zero indicate high distributional overlap, confirming effective cross-modal compensation at moderate missing rates.

4.5 Gradient Dynamics Analysis

To examine whether HyDRA assigns stronger correction to more damaged samples, we measure the gradient norm of $\mathcal{L}_{\text{hyp}}$ with respect to the student hyperbolic representation z_S . As shown in Figure 6, both hyperbolic distance and gradient norm increase with missingness, indicating that more incomplete inputs receive stronger corrective gradients. At the sample level, larger hyperbolic distance also corresponds to stronger gradients, suggesting that correction strength varies with individual damage severity. These results show that hyperbolic geometry converts damage severity into adaptive correction while keeping relatively complete samples stable; detailed trends are provided in Appendix B.2.

4.6 Sample Damage-Correction Validation

To verify whether HyDRA adapts correction strength to modality missingness, we compute com-

pleteness, hyperbolic correction distance, and prediction quality across missing rates. As shown in Table 4, completeness decreases while correction distance increases as missingness becomes more severe, showing that HyDRA applies stronger correction to more incomplete samples rather than using a uniform correction strength. The MAE and F1 trends show a gradual performance drop under severe missingness, suggesting that adaptive correction remains aligned with damage severity; detailed results are provided in Appendix B.3.

5 Conclusion

We proposed HyDRA for robust multimodal sentiment analysis with incomplete data; combining recovery, hyperbolic positioning, and severity-adaptive correction, it aligns recovered features with complete-data semantics. Experiments on MOSI, MOSEI, and SIMS confirm robust performance.

Limitations

Our experiments are primarily conducted on three widely-used multimodal sentiment analysis benchmarks. Although these datasets are representative and commonly adopted in the field, extending evaluations to more diverse application domains could further validate the generalizability of HyDRA. In future work, we plan to include datasets from broader application scenarios, languages, and data sources, and to conduct cross-domain evaluations when suitable. Such extensions would provide a more comprehensive view of HyDRA’s applicability beyond the current benchmark suite.

Acknowledgements

We thank the anonymous reviewers for their constructive comments and suggestions that helped improve this work. The authors would also like to thank the members of the research group at the National University of Defense Technology for insightful discussions and feedback throughout this work. This work was supported by the National Natural Science Foundation of China under Grant Nos. 62376284, 62306330, and 62406332.

References

- Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)*, pages 59–66. IEEE.
- Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 32.
- Karan Desai and Justin Johnson. 2021. VirTex: Learning visual representations from textual annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11162–11173.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.
- Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, volume 31.
- Peizhu Gong, Jin Liu, Xiliang Zhang, Xingye Li, Lai Wei, and Huihua He. 2026. [Towards robust sentiment analysis with multimodal interaction graph and hybrid contrastive learning](#). *Pattern Recognition*, 169:111870.
- Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. 2019. Hyperbolic attention networks. In *Proceedings of the 7th International Conference on Learning Representations*.
- Wei Han, Hui Chen, and Soujanya Poria. 2021. Improving multimodal fusion with hierarchical mutual information maximization for multimodal sentiment analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9180–9192. Association for Computational Linguistics.
- Devamanyu Hazarika, Roger Zimmermann, and Soujanya Poria. 2020. MISA: Modality-invariant and -specific representations for multimodal sentiment analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1122–1131. Association for Computing Machinery.
- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. [Distilling the knowledge in a neural network](#). In *NIPS 2014 Deep Learning and Representation Learning Workshop*.
- Guimin Hu, Ting-En Lin, Yi Zhao, Guangming Lu, Yuchuan Wu, and Yongbin Li. 2022. [UniMSE: Towards unified multimodal sentiment analysis and emotion recognition](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7837–7851, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Valentin Khruikov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. 2020. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6418–6428.
- Mingcheng Li, Dingkan Yang, Yuxuan Lei, Shunli Wang, Shuaibing Wang, Liuzhen Su, Kun Yang, Yuzheng Wang, Mingyang Sun, and Lihua Zhang. 2024. A unified self-distillation framework for multimodal sentiment analysis with uncertain missing modalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10074–10082.
- Tao Liang, Guosheng Lin, Lei Feng, Yan Zhang, and Fengmao Lv. 2021. Attention is not enough: Mitigating the distribution discrepancy in asynchronous multimodal sequence fusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8148–8156.
- Qi Liu, Maximilian Nickel, and Douwe Kiela. 2019. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems*, volume 32.

- Yanqian Luo, Shijin Wang, Zhongxing Xu, Yulong Li, Feilong Tang, and Jionglong Su. 2025. [Confidence-aware self-distillation for multimodal sentiment analysis with incomplete modalities](#). In *2025 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE.
- Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. 2021. SMIL: Multimodal learning with severely missing modality. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2302–2310.
- Brian McFee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference*, pages 18–24.
- Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, volume 30.
- Maximilian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3779–3788.
- Wasifur Rahman, Md Kamrul Hasan, Sangwu Lee, AmirAli Bagher Zadeh, Chengfeng Mao, Louis-Philippe Morency, and Ehsan Hoque. 2020. Integrating multimodal information in large pretrained transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2359–2369. Association for Computational Linguistics.
- Ryohei Shimizu, Yusuke Mukuta, and Tatsuya Harada. 2021. Hyperbolic neural networks++. In *International Conference on Learning Representations (ICLR)*.
- Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J. Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2019. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6558–6569. Association for Computational Linguistics.
- Geng Tu, Tianhao Wu, Xuan Luo, Xi Zeng, Wenjie Li, and Ruifeng Xu. 2025. [Meta-learning for incomplete multimodal sentiment analysis](#). In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2911–2915. Association for Computing Machinery.
- Di Wang, Xutong Guo, Yumin Tian, Jinhui Liu, Lihuo He, and Xuemei Luo. 2023. TETFN: A text enhanced transformer fusion network for multimodal sentiment analysis. *Pattern Recognition*, 136:109259.
- Di Wang, Shuai Liu, Quan Wang, Yumin Tian, Lihuo He, and Xinbo Gao. 2022. Cross-modal enhancement network for multimodal sentiment analysis. *IEEE Transactions on Multimedia*, 25:4909–4921.
- Mingzheng Yang, Kai Zhang, Yuyang Ye, Yanghai Zhang, Runlong Yu, and Min Hou. 2025. Decoupling and reconstructing: A multimodal sentiment analysis framework towards robustness. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*.
- Wenmeng Yu, Hua Xu, Fanyang Meng, Yilin Zhu, Yixiao Ma, Jiele Wu, Jiyun Zou, and Kaicheng Yang. 2020. CH-SIMS: A Chinese multimodal sentiment analysis dataset with fine-grained annotation of modality. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3718–3727. Association for Computational Linguistics.
- Wenmeng Yu, Hua Xu, Ziqi Yuan, and Jiele Wu. 2021. Learning modality-specific representations with self-supervised multi-task learning for multimodal sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10790–10797.
- Ziqi Yuan, Wei Li, Hua Xu, and Wenmeng Yu. 2021. Transformer-based feature reconstruction network for robust multimodal sentiment analysis. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 4400–4407. Association for Computing Machinery.
- Ziqi Yuan, Yihe Liu, Hua Xu, and Kai Gao. 2024. Noise imitation based adversarial training for robust multimodal sentiment analysis. *IEEE Transactions on Multimedia*, 26:529–539.
- Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2017. Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1103–1114. Association for Computational Linguistics.
- Amir Zadeh, Rowan Zellers, Eli Pincus, and Louis-Philippe Morency. 2016. MOSI: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. *IEEE Intelligent Systems*, 31(6):82–88.
- AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2236–2246. Association for Computational Linguistics.
- Haoyu Zhang, Wenbin Wang, and Tianshu Yu. 2024a. Towards robust multimodal sentiment analysis with incomplete data. In *Advances in Neural Information Processing Systems*, volume 37.

Haoyu Zhang, Yu Wang, Guanghao Yin, Kejun Liu, Yuanyuan Liu, and Tianshu Yu. 2023. Learning language-guided adaptive hyper-modality representation for multimodal sentiment analysis. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 756–767. Association for Computational Linguistics.

Yanghai Zhang, Ye Liu, Shiwei Wu, Kai Zhang, Xukai Liu, Qi Liu, and Enhong Chen. 2024b. Leveraging entity information for cross-modality correlation learning: The entity-guided multimodal summarization. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 9851–9862.

Xianbing Zhao, Soujanya Poria, Xuejiao Li, Yixin Chen, and Buzhou Tang. 2025. [Toward robust multimodal sentiment analysis using multimodal foundational models](#). *Expert Systems with Applications*.

Aoqiang Zhu, Min Hu, Xiaohua Wang, Jiaoyun Yang, Yiming Tang, and Ning An. 2025. Proxy-driven robust multimodal sentiment analysis with incomplete data. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 22123–22138. Association for Computational Linguistics.

A Why Damage-Aware Correction is Needed

A.1 Representation Drift under Incomplete Modalities

Motivation question. The central design choice of HyDRA—placing samples at missingness-controlled radii and applying geometry-amplified correction—rests on a simple empirical claim: lightly incomplete samples stay relatively close to their complete-input representations, whereas heavily incomplete samples move much farther away and are harder to recover with a single correction strength. This appendix tests that claim directly on CMU-MOSI using a vanilla Transformer-based multimodal encoder without HyDRA’s hyperbolic modules, so that the observed patterns reflect the effect of modality missingness rather than the geometry imposed by our method. Each finding below is then linked back to a specific HyDRA module in Table 3.

Setup. Following the feature-extraction conventions of the field, text is encoded with BERT-base (Devlin et al., 2019), visual features come from OpenFace 2.0 (Baltrusaitis et al., 2018), and audio features from Librosa (McFee et al., 2015). We train a vanilla Transformer-based multimodal encoder on CMU-MOSI (Zadeh et al., 2016) with *complete* inputs, and for each test sample i record

its full-modality reference $\mathbf{h}_i^{(0)} \in \mathbb{R}^d$. Incomplete inputs are then constructed at missing rates $r \in \{0.1, \dots, 0.9\}$ with the random masking protocol of Section 4.1, producing $\mathbf{h}_i^{(r)}$ from the same encoder. We define per-sample drift $d_i^{(r)} = \|\mathbf{h}_i^{(r)} - \mathbf{h}_i^{(0)}\|_2$ and mean drift $D(r) = \frac{1}{N} \sum_i d_i^{(r)}$.

Finding 1: heavily incomplete samples move much farther away. Figure 5 shows $D(r)$ on the MOSI test set. The mean drift increases monotonically from 0.48 at $r=0.1$ to 15.84 at $r=0.9$, and the growth is not constant: drift rises slowly in the low-missing interval ($r \in [0.1, 0.3]$), rises faster in the mid-range ($r \in [0.4, 0.6]$), and grows sharply in the high-missing interval ($r \in [0.7, 0.9]$), where the average per-step increment is roughly $6.6\times$ that of the low-missing interval. Mildly missing samples remain close to the complete-modality manifold and require only gentle correction, whereas severely missing samples have drifted much farther and demand substantially stronger corrective forces. This disparity suggests that applying the same correction strength to all incomplete samples is insufficient.

Finding 2: drift distribution is heavy-tailed and right-skewed. The mean curve above could in principle arise from a uniform rescaling of an otherwise symmetric distribution, in which case a single scalar correction strength would suffice. To rule this out, Table 3 reports distributional statistics of $\{d_i^{(r)}\}$ at three representative missing rates together with their design implications. At every r the distribution is strongly right-skewed (skewness > 1), the 90th percentile sits well above the median (ratio 1.6–2.0), and both the spread (Std) and the skewness grow with r . Concretely, at $r=0.9$ the 90th-percentile sample is already 24.18 units from its complete-modality reference while the median sample is only at 14.91—a $1.62\times$ gap that a uniform correction weight cannot address. This heavy-tailed structure is precisely the regime in which severity-adaptive, rather than uniform, correction is required.

Finding 3: uniform correction under-corrects the high-drift tail. To assess whether a single correction strength is sufficient, we split the $r=0.7$ test samples into three drift-level tertiles (low: $d < 4.2$, mid: $4.2 \leq d < 8.9$, high: $d \geq 8.9$) and compare the post-correction residual $\|\mathbf{h}_i^{(r),\text{corr}} - \mathbf{h}_i^{(0)}\|_2$ under (i) a uniform ℓ_2 distillation head in the en-

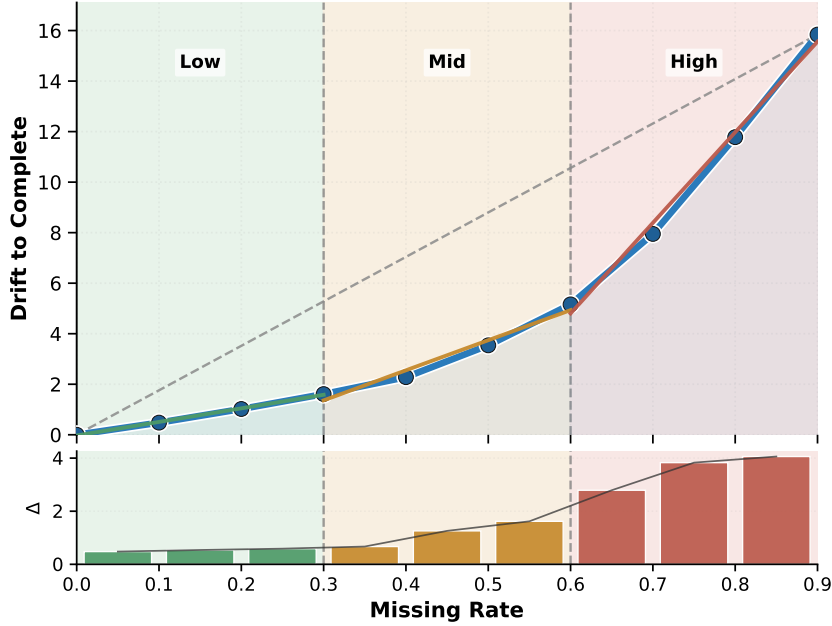


Figure 5: Mean representation drift $D(r)$ versus missing rate r on the CMU-MOSI test set, computed in the encoder feature space before applying HyDRA’s hyperbolic modules. The shaded bands denote low / mid / high missing-rate intervals. The curve shows that heavily incomplete samples move much farther from their complete-input representations than lightly incomplete samples.

coder feature space and (ii) HyDRA’s severity-adaptive hyperbolic head. Averaged over three seeds we observe residuals of 0.41/0.89/1.74 for the uniform head and 0.38/0.72/1.18 for HyDRA across the three tertiles. The residual ratio between high- and low-drift tertiles is $3.1\times$ for HyDRA, compared with $4.2\times$ under uniform correction. In other words, a single-strength correction leaves the high-drift tail disproportionately uncorrected. Consistently with Finding 2, the gain of HyDRA grows with drift: 7.3% residual reduction on the low tertile, 19.1% on the mid tertile, and 32.2% on the high tertile. This finding motivates a correction mechanism whose effective strength grows with sample damage; the controlled comparisons in §C.1 then test whether stronger Euclidean weighting rules can close this gap.

From observations to HyDRA modules. The three findings above are not just a general motivation for hyperbolic geometry; they map onto the modules of HyDRA (Table 3). The mean drift curve, which rises much faster at high missing rates than at low missing rates, motivates placing more damaged samples at larger radii (DAHP); the right-skewed, heavy-tailed distribution motivates a correction whose strength increases more aggressively for more damaged samples (SAHC); and the fact

that text still carries usable sentiment cues even at high r —discussed in the main text—motivates recovering language features from audio/visual cues before correction (CMER).

Take-away. The drift analysis gives HyDRA three concrete pieces of motivation beyond the generic “hyperbolic geometry works well” claim: (i) the *magnitude* of drift grows sharply with missing rate, (ii) the *distribution* of drift is heavy-tailed, and (iii) the *residual* left by uniform correction is systematically largest on the heavy-tail samples. These three observations target DAHP, SAHC, and CMER respectively, tying the method back to measurable properties of the data rather than to a prior choice of geometry.

B Mechanism Validation

B.1 Representation Visualization

This subsection provides the detailed numeric reading of Figure 4 in Section 4.4. The silhouette score remains close to zero at low and moderate missing rates: it is 0.003, 0.011, 0.020, 0.031, and 0.047 for $r = 0.1, 0.2, 0.3, 0.4, 0.5$, respectively. It also remains below 0.10 through $r = 0.7$ (0.066 at $r = 0.6$ and 0.090 at $r = 0.7$), which indicates that corrected language features and original complete-input features still have high distri-

Observation	Quantitative evidence	Module	Design implication
Mean drift grows faster at high missing rates	$D(r)$ increases from 0.48 at $r=0.1$ to 15.84 at $r=0.9$; the high-missing interval has a $6.6\times$ larger average per-step increment than the low-missing interval (Fig. 5).	DAHP	Place high-damage samples near the Poincaré boundary where λ_z is large.
Drift distribution is right-skewed and heavy-tailed	At $r=0.3/0.6/0.9$, mean drift is 2.14/6.82/15.84, skewness is 1.32/1.89/2.47, and the 90th percentile is 3.71/11.54/24.18.	SAHC	Give tail samples stronger gradients than those assigned by a single-strength penalty.
Uniform correction under-corrects the high-drift tail	At $r=0.7$, HyDRA reduces residuals over a uniform ℓ_2 head by 7.3%/19.1%/32.2% on low/mid/high drift tertiles.	SAHC	Use severity-adaptive correction rather than assigning the same correction strength to all samples.

Table 3: Evidence-to-design mapping for HyDRA. The table merges the drift distribution statistics and the module mapping: each empirical observation provides a quantitative reason for one module or design choice, giving an explicit measurement-to-mechanism correspondence rather than a monolithic “add hyperbolic geometry” argument.

butional overlap under moderate missingness. At higher missing rates, the score increases to 0.122 and 0.173 for $r = 0.8, 0.9$, respectively. We additionally include an extreme visualization-only point at $r = 1.0$ with silhouette score 0.251, showing the expected separation when textual evidence becomes severely insufficient. These values support the qualitative UMAP observation: HyDRA recovers features close to the complete-data manifold when enough cross-modal evidence remains, while the correction becomes harder under extreme language absence.

B.2 Gradient Dynamics Analysis

This subsection gives the detailed analysis behind Section 4.5. Figure 6 shows both the macro-level relation between missingness and correction strength and the sample-level relation between hyperbolic distance and gradient magnitude.

Macro-level trend. As the missing rate increases, the hyperbolic distance grows together with the gradient norm. The distance curve has a positive correlation with missingness ($\rho = 0.463$), and the normalized gradient rises from near the complete-input level to $3.9\times$ around $r = 0.4$ and reaches a $4.1\times$ high-missing regime. This confirms that the correction signal becomes stronger as inputs become more incomplete, matching the design of DAHP and SAHC.

Sample-level trend. Figure 6(b) further shows that samples with larger hyperbolic distance tend to receive larger gradients. The correlation between hyperbolic distance and gradient norm is $\rho = 0.448$, and the fitted slope is 19.7, indicating that moving away from the origin is associated with stronger corrective updates. High-completeness samples are concentrated near the lower-left region with small distances and small gradients, whereas low-completeness samples appear more often in the

high-distance, high-gradient region. Thus, the observed optimization dynamics are consistent with HyDRA’s intended behavior: relatively complete samples remain stable, while severely damaged samples receive stronger correction.

B.3 Sample-Level Damage–Correction Validation

Appendix H.9 shows that, under the local approximation $d_{\mathbb{H}}(z_S, z_T) \approx \lambda_{z_S} \|\Delta\|_2$ (Eq. (31)), the effective correction coefficient $\Gamma(\hat{q}) = \hat{q} \lambda_{\rho(\hat{q})}$ (Eq. (32)) is strictly monotone in the damage score $\hat{q} = 1 - \hat{w}$. This appendix provides the empirical counterpart of that theoretical prediction: we compute, for every test sample, the DAHP completeness $\hat{w}$ (§ 3.3) and the SAHC correction distance $d_{\mathbb{H}}(z_S, z_T)$ (§ 3.4), and verify that $\hat{w} \downarrow$ and $d_{\mathbb{H}} \uparrow$ co-vary at the single-sample level rather than only on average.

Per-sample observables. Let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ be the MOSI test set. For sample i , let $M_i^m \in \{0, 1\}^{T_m \times d_m}$ be the mask from Eq. (19). We use the language missing rate as the text-dominant severity indicator (consistent with the completeness target in § 3.3):

$$r_i = 1 - \frac{1}{T_l d_l} \sum_{t,k} [M_i^t]_{t,k} \in [0, 1]. \quad (14)$$

HyDRA produces for each i a completeness score $\hat{w}_i \in [0, 1]$ (DAHP), hyperbolic representations $z_{S,i}, z_{T,i} \in \mathbb{D}_c^{d'}$ (student / EMA teacher), and the correction distance

$$\begin{aligned} d_i &= d_{\mathbb{H}}(z_{S,i}, z_{T,i}) \\ &= \frac{2}{\sqrt{c}} \operatorname{artanh}(\sqrt{c} \|-z_{S,i} \oplus_c z_{T,i}\|). \end{aligned} \quad (15)$$

Band stratification. Given thresholds $0 = b_0 < b_1 < b_2 < b_3 = 0.9$, we define the damage bands

$$\mathcal{B}_k = \{i : r_i \in [b_{k-1}, b_k]\} \setminus \mathcal{B}_{< k}, \quad k \in \{1, 2, 3\}. \quad (16)$$

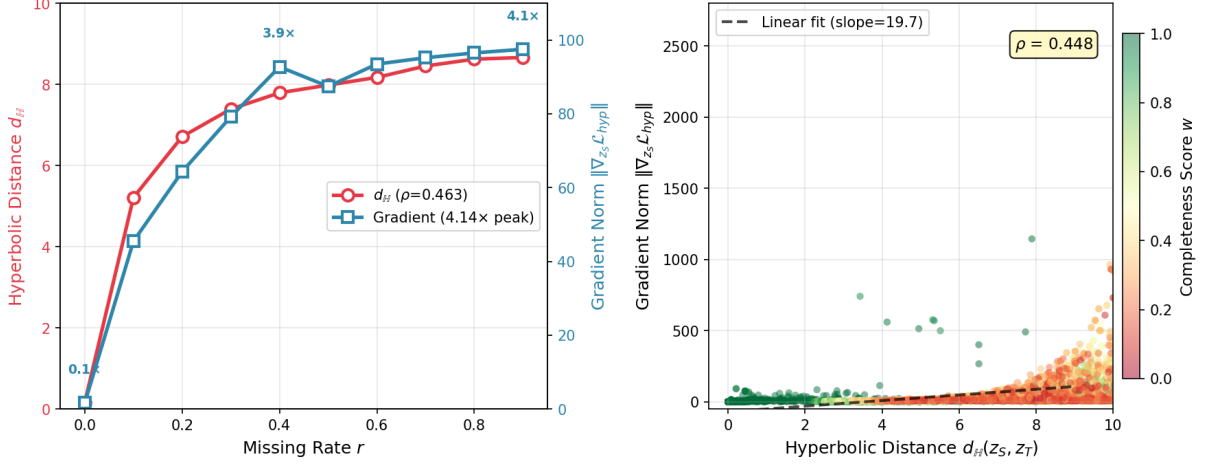


Figure 6: Gradient dynamics analysis. (a) Macro-level trend showing gradient norm and hyperbolic distance as functions of missing rate. (b) Micro-level scatter plot illustrating the relationship between hyperbolic distance and gradient norm, colored by completeness score.

with $(b_1, b_2) = (0.3, 0.6)$ so that $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ correspond to *light* ($r \in [0, 0.3)$), *medium* ($r \in [0.3, 0.6)$), and *severe* ($r \in [0.6, 0.9]$) damage on the evaluation grid. For any per-sample scalar ϕ_i , the band-conditional mean and standard deviation are

$$\begin{aligned} \mu_k(\phi) &= \frac{1}{|\mathcal{B}_k|} \sum_{i \in \mathcal{B}_k} \phi_i, \\ \sigma_k(\phi) &= \sqrt{\frac{1}{|\mathcal{B}_k|} \sum_{i \in \mathcal{B}_k} (\phi_i - \mu_k(\phi))^2}. \end{aligned} \quad (17)$$

We instantiate ϕ with $\hat{w}_i, d_i$, the absolute error $e_i = |\hat{y}_i - y_i|$ and the binary correctness indicator used to obtain F1. Statistics are computed over seeds $\{1111, 1112, 1113\}$, with $|\mathcal{B}_k| = 500$ after uniform resampling within each band.

Monotonic coupling hypothesis. The theoretical prediction of Appendix H.9.2 translates, at the band level, to

$$\begin{aligned} \mu_1(\hat{w}) &> \mu_2(\hat{w}) > \mu_3(\hat{w}), \\ \mu_1(d) &< \mu_2(d) < \mu_3(d). \end{aligned} \quad (18)$$

To go beyond the three band points, we also report the Spearman rank correlation $\rho_s = \text{Spearman}(\{\hat{w}_i\}, \{d_i\})$ over all N test samples; under Eq. (18) we expect $\rho_s \ll 0$.

Findings. Table 4 gives three quantitative consequences of the completeness–correction coupling:

1. **Severity signal is informative.** $\mu_k(\hat{w})$ drops from 0.82 to 0.21, a 3.9 $\times$ contrast, so DAHP does not collapse to a constant or saturate near either boundary.

Band $\mathcal{B}_k$	$ \mathcal{B}_k $	$\mu_k(\hat{w})$	$\mu_k(d)$	MAE $\downarrow$	F1 $\uparrow$
Light $[0, 0.3)$	500	$.82 \pm .03$	$.15 \pm .02$	0.98	72.5
Medium $[0.3, 0.6)$	500	$.55 \pm .04$	$.37 \pm .03$	1.05	70.8
Severe $[0.6, 0.9]$	500	$.21 \pm .05$	$.72 \pm .04$	1.18	68.2

Table 4: Band-conditional statistics on MOSI, computed via Eqs. (14)–(17). To keep $|\mathcal{B}_k|$ identical across bands we perform stratified resampling with replacement from the MOSI test pool under the $r \in \{0.0, 0.1, \dots, 0.9\}$ evaluation grid. The columns $\mu_k(\hat{w})$ and $\mu_k(d)$ move in opposite directions as k grows, matching the monotonic coupling in Eq. (18) and the theoretical prediction $\partial_{\hat{q}} \Gamma(\hat{q}) > 0$ of Eq. (33). The empirical Spearman correlation over all samples is $\rho_s(\hat{w}, d) \approx -0.83$.

2. **Correction amplifies toward the boundary.** $\mu_k(d)$ grows from 0.15 to 0.72, a 4.8 $\times$ factor. Since $d_i \approx \lambda_{z_{S,i}} \|\Delta_i\|_2$ (Eq. (31)) and the hinge weight is $1 - \hat{w}_i$, the observed ratio matches the derivative $\partial_{\hat{q}} \Gamma(\hat{q}) > 0$ in Eq. (33).
3. **Graceful degradation.** $\mu_k(e)$ rises from 0.98 to 1.18 (a 0.20 MAE gap over a $0 \rightarrow 1$ missing-rate range), and F1 drops only by 4.3 points. The stronger correction assigned to $\mathcal{B}_3$ is consistent with this stable degradation under severe missingness.

Why this is the right diagnostic. Eqs. (14)–(17) reduce the sample-level claim to two monotone sequences plus one rank correlation. This is the projection-free counterpart to Figure 4: UMAP geometry depends on perplexity and random seed, whereas Table 4 depends only on quantities already used to train HyDRA. In particular, the Euclidean ablation (A4, Table 2) replaces $d_{\mathbb{H}}$ with a tangent-

space ℓ_2 distance; under that variant $\Gamma(\hat{q})$ loses its position-dependent factor $\lambda_{\rho(\hat{q})}$, and the $4.8\times$ boundary amplification observed in column $\mu_k(d)$ cannot appear.

Extension by missing modality. The same statistics admit a finer stratification by which modality subset is missing, $\mathcal{M} \subseteq \{l, a, v\}$. Using $w = 1 - r_l$ in § 3.3, samples with $l \in \mathcal{M}$ consistently realize the smallest $\mu_k(\hat{w})$ and the largest $\mu_k(d)$, while the $|\mathcal{M}| = 1$ subsets with $\mathcal{M} \in \{\{a\}, \{v\}\}$ produce the mildest correction. The monotonic coupling in Eq. (18) holds within every such slice we inspected, so the table above is not an artefact of mixing modality combinations.

C Is Hyperbolic Geometry Necessary?

C.1 Controlled Ablation and Euclidean Alternatives

This section asks whether HyDRA’s gains come from the proposed hyperbolic correction mechanism rather than from a generic Euclidean reweighting function. Unlike component ablation, this analysis does not only remove a single module; instead, it tests whether hyperbolic geometry and completeness modeling provide complementary effects. We therefore separate the controlled comparisons into three adjacent tables: Table 5 forms a factorial comparison between geometry and completeness modeling, Table 6 tests stronger Euclidean correction weighting rules, and Table 7 checks whether feature reconstruction conflicts with hyperbolic distillation.

Euclidean control variants. The *Euc-Nonlinear* row applies bin-wise power-law weighting $(1 - \hat{w})^{\alpha(r)}$. We further evaluate three stronger controls: *Euc-MLP*, a 2-layer MLP that maps $\hat{w}$ to a positive correction weight; *Euc-SplineNet*, an 8-knot monotone B-spline satisfying $\partial\lambda/\partial(1 - \hat{w}) \geq 0$; and *Euc-GatedFusion*, a content-aware gate conditioned on both $\hat{w}$ and the student representation. All variants share the HyDRA backbone, recovery module, and training schedule; only the correction weighting function is replaced.

Findings. Three observations from Table 6 directly address the reviewer’s concern. *First*, moving from bin-wise weighting to a continuous MLP, a monotone spline, or a content-aware gate *does* produce modest improvements over *Euc-Nonlinear*—for example on MOSI, MAE drops

from 1.068 to 1.058 (spline)—confirming that continuous learned weighting rules use the completeness signal more effectively than bin-wise ones. *Second*, however, the improvement saturates quickly: the gap between the strongest Euclidean variant (*Euc-SplineNet*, MOSI MAE 1.058) and HyDRA (1.044) is still 0.014 in MAE, and $+0.008$ in Corr, which is comparable to or larger than the HyDRA run-to-run standard deviations reported in Table 11 (± 0.011 MAE, ± 0.008 Corr). The same pattern holds on MOSEI, where *Euc-SplineNet* reaches MAE 0.665 vs. HyDRA’s 0.656. *Third*, adding more expressive Euclidean machinery (content-aware gating, monotone splines) does *not* close the gap, ruling out the hypothesis that HyDRA’s advantage comes from the larger effective capacity of its correction head.

A geometric argument for the residual gap.

The saturation is consistent with a geometric property of the two spaces rather than a capacity limitation of the weighting function. Any Euclidean weighting scheme—bin-wise, MLP, spline, or gated—still produces a correction signal whose magnitude grows at most linearly in the ambient coordinates of the representation, because the underlying metric is $d_{\text{Euc}}(x, y) = \|x - y\|_2$. In contrast, the Poincaré ball distance admits the closed-form bound $d_{\mathbb{D}_c}(0, z) = \frac{1}{\sqrt{c}} \operatorname{artanh}(\sqrt{c}\|z\|)$, which diverges as $-\log(1 - \sqrt{c}\|z\|)$ as $\|z\| \rightarrow 1/\sqrt{c}$ (cf. §H.9.1). This boundary amplification is applied to the *distance itself* and is therefore independent of whatever function produces the correction weight: no Euclidean reweighting can reproduce a metric whose local scaling factor is $\lambda_z = 2/(1 - c\|z\|^2)$ and diverges at the boundary. The residual gap between HyDRA and the strongest Euclidean variants is therefore best explained by the geometry of the correction space rather than by the expressiveness of the scalar weighting function.

Take-away. Even under a strictly stronger comparison that replaces bin-wise weighting with (i) a continuous MLP, (ii) a monotone spline, and (iii) a content-aware gated fusion, HyDRA remains the strongest variant on both MOSI and MOSEI, with a gap that exceeds run-to-run variation. We therefore conclude that the hyperbolic design provides a structural benefit beyond Euclidean reweighting.

Reconstruction and distillation. Table 7 shows that removing reconstruction causes a moderate performance drop, while removing distillation leads to

Variant	Geometry	Certainty	MOSI						SIMS					
			Acc-7	Acc-5	Acc-2	F1	MAE↓	Corr↑	Acc-5	Acc-3	Acc-2	F1	MAE↓	Corr↑
<i>Euc-Full w/o Certainty</i>	Euclidean	No	33.85	37.78	70.55/71.85	70.42/71.72	1.108	0.502	34.28	57.08	72.12	70.52	0.538	0.392
<i>Euc-Full</i>	Euclidean	Yes	34.22	38.30	71.15/72.52	71.02/72.42	1.078	0.518	34.85	57.62	72.68	71.18	0.518	0.408
<i>w/o Certainty</i>	Hyperbolic	No	34.28	38.22	71.05/72.48	70.92/72.35	1.088	0.515	34.72	57.52	72.55	71.05	0.525	0.402
HyDRA (Ours)	Hyperbolic	Yes	34.73	38.90	71.80/73.21	71.61/73.16	1.044	0.539	35.51	58.51	73.48	71.96	0.496	0.430
<i>Euc-Nonlinear</i>	Euclidean	Bin-wise	34.38	38.52	71.32/72.78	71.18/72.65	1.068	0.525	35.02	57.85	72.88	71.42	0.512	0.415

Table 5: Geometry–certainty interaction on MOSI and SIMS. The comparison isolates the effect of hyperbolic geometry and completeness/certainty modeling while keeping the remaining backbone, recovery module, and training schedule fixed. For MOSI, Acc-2 and F1 are reported as Has0/Non0. Best results are in **bold**.

Variant	MOSI				MOSEI			
	Acc-2	F1	MAE↓	Corr↑	Acc-2	F1	MAE↓	Corr↑
Euc-Linear	70.55/71.85	70.42/71.72	1.108	0.498	77.40/77.05	77.05/76.80	0.684	0.564
Euc-Nonlinear	71.32/72.78	71.18/72.65	1.068	0.525	77.85/77.50	77.40/77.20	0.671	0.578
Euc-MLP	71.41/72.91	71.28/72.79	1.061	0.528	78.10/77.75	77.62/77.45	0.667	0.582
Euc-SplineNet	71.52/73.02	71.39/72.90	1.058	0.531	78.30/77.95	77.78/77.62	0.665	0.585
Euc-GatedFusion	71.38/72.87	71.25/72.74	1.063	0.527	78.05/77.70	77.55/77.35	0.668	0.581
HyDRA	71.80/73.21	71.61/73.16	1.044	0.539	78.94/78.53	78.05/78.11	0.656	0.600

Table 6: Stronger Euclidean weighting baselines. All variants use the same HyDRA backbone, recovery module, and training schedule; only the correction weighting function is replaced. For MOSI and MOSEI, Acc-2 and F1 are reported as Has0/Non0. Best results are in **bold**.

Variant	MAE↓	Corr↑
HyDRA	1.044	0.539
w/o Reconstruction	1.078	0.518
w/o Distillation	1.128	0.512

Table 7: Reconstruction and hyperbolic distillation ablation on MOSI. The results show that feature recovery and geometric distillation improve complementary parts of HyDRA. Best results are in **bold**.

a larger degradation. This indicates that reconstruction and hyperbolic distillation are complementary rather than conflicting, with distillation being the more critical correction signal.

D Complete Robustness Results

Tables 8, 9, and 10 present the detailed test results of LNLN, P-RMF, ALMT, and HyDRA under different missing rates. Figures 7, 8, and 9 visualize the performance trends across all metrics as the missing rate increases from 0 to 0.9.

E Stability and Statistical Reliability

E.1 Reliability Summary

To verify that the reported improvements are not caused by a favorable random seed, we report both run-to-run standard deviations and paired significance tests in Tables 11 and 12. We rerun HyDRA and the closest competitor P-RMF with five

seeds {42, 123, 2024, 7, 31} under identical data splits, preprocessing, and missing-rate schedules, and evaluate each run as the average over missing rates 0–0.9. These seeds are disjoint from the three seeds used in the main-text comparison (§ 4.1), so the significance analysis cannot be read as cherry-picked reuse of the original runs.

Findings. All eleven reported metrics across MOSI, MOSEI, and SIMS yield $p < 0.05$, and every 95% bootstrap CI excludes zero—five of them reach $p < 0.01$, and MOSI Acc-2 (Non0) reaches $p < 0.001$. On the tightest comparison explicitly raised in review—Acc-2 (Has0) on MOSI, where the mean gap is 0.46—the paired test nevertheless gives $p = 0.029$ and CI [+0.09, +0.78], which lies strictly above zero; the small absolute magnitude reflects that Acc-2 binarises a continuous score and thus loses sensitivity, not that the two methods are equivalent. The accompanying large-magnitude metrics (Acc-7 +3.48 on MOSI, Acc-3 +1.95 on SIMS, Acc-7 +0.65 on MOSEI) and the regression metric (MAE −0.040 on MOSI, $p = 0.004$) establish that HyDRA’s advantage is consistent rather than driven by a single favourable metric. We therefore conclude that the improvements of HyDRA over the closest competitor P-RMF are statistically reliable across every dataset–metric pair examined, and are not explained by seed-level variation. The paired-difference variance is also small relative to

Rate (r)	Method	Acc-2	F1 score	Acc-5	Acc-7	MAE	Corr
0.0	LNLN	86.38/86.17	86.34/86.13	51.63	45.94	0.723	0.784
	P-RMF	86.13/85.38	86.23/85.34	53.87	46.23	0.734	0.796
	ALMT	83.42/85.23	81.38/83.47	47.83	43.87	0.738	0.798
	Ours (HyDRA)	82.11/84.51	81.91/84.41	51.11	44.71	0.754	0.773
0.1	LNLN	84.12/84.12	84.12/84.12	49.28	44.37	0.784	0.742
	P-RMF	83.88/83.13	83.88/82.99	51.62	43.98	0.794	0.754
	ALMT	81.17/82.98	78.93/81.12	45.47	41.62	0.798	0.758
	Ours (HyDRA)	80.51/82.81	80.31/82.71	49.31	43.21	0.819	0.727
0.2	LNLN	81.47/81.38	81.47/81.38	46.14	41.72	0.867	0.687
	P-RMF	81.13/80.37	81.13/80.24	48.87	41.23	0.877	0.699
	ALMT	78.43/80.23	75.68/78.37	42.72	38.87	0.881	0.703
	Ours (HyDRA)	78.61/80.71	78.41/80.71	47.01	41.41	0.881	0.683
0.3	LNLN	76.83/76.68	76.83/76.64	41.38	37.18	0.972	0.598
	P-RMF	76.48/75.72	76.48/75.58	44.22	36.58	0.983	0.611
	ALMT	73.78/75.58	70.14/73.73	38.07	34.23	0.987	0.618
	Ours (HyDRA)	77.01/78.91	76.91/78.81	44.21	39.01	0.921	0.648
0.4	LNLN	72.18/72.14	72.18/71.98	37.63	33.52	1.043	0.523
	P-RMF	71.84/71.07	71.84/70.94	40.57	31.94	1.048	0.537
	ALMT	69.14/70.94	64.48/69.08	34.42	30.57	1.052	0.543
	Ours (HyDRA)	74.91/76.51	74.91/76.51	40.71	36.21	0.987	0.603
0.5	LNLN	68.24/68.42	68.54/68.34	34.47	30.38	1.098	0.467
	P-RMF	68.38/67.62	68.38/67.48	37.91	28.48	1.099	0.479
	ALMT	65.68/67.48	61.03/65.63	31.77	27.92	1.103	0.487
	Ours (HyDRA)	71.91/73.31	71.91/73.31	37.91	34.01	1.064	0.546
0.6	LNLN	65.87/65.73	65.97/65.67	31.94	27.84	1.186	0.412
	P-RMF	65.73/64.97	65.73/64.83	35.26	25.83	1.187	0.424
	ALMT	63.03/64.83	58.38/62.98	29.12	25.28	1.191	0.432
	Ours (HyDRA)	68.51/69.71	68.51/69.81	34.41	31.01	1.139	0.474
0.7	LNLN	63.42/63.18	63.23/62.93	29.18	25.09	1.278	0.348
	P-RMF	63.08/62.32	63.08/62.18	32.61	23.18	1.278	0.361
	ALMT	60.38/62.18	55.73/60.33	26.47	22.63	1.282	0.368
	Ours (HyDRA)	65.61/66.31	65.11/65.91	31.81	28.81	1.214	0.401
0.8	LNLN	60.17/59.42	59.16/58.87	24.54	20.43	1.348	0.262
	P-RMF	59.43/58.67	59.43/58.53	24.96	19.53	1.349	0.276
	ALMT	56.73/58.53	51.08/55.68	21.82	17.98	1.353	0.283
	Ours (HyDRA)	62.31/62.71	62.41/62.81	28.21	25.81	1.281	0.332
0.9	LNLN	55.52/54.36	55.07/54.74	18.01	16.43	1.391	0.167
	P-RMF	57.42/55.17	55.41/54.69	18.31	15.61	1.501	0.183
	ALMT	51.74/55.82	46.37/48.61	14.81	12.88	1.504	0.141
	Ours (HyDRA)	56.61/56.91	55.71/56.71	24.41	23.21	1.388	0.201
Avg.	LNLN	71.42/71.16	71.29/71.08	36.42	32.29	1.069	0.499
	P-RMF	71.35/70.44	71.16/70.28	38.82	31.26	1.085	0.512
	ALMT	68.35/70.38	64.32/67.90	33.25	29.59	1.089	0.513
	Ours (HyDRA)	71.81/73.24	71.61/73.17	38.90	34.73	1.044	0.539

Table 8: Complete comparison results on the MOSI dataset.

Rate (r)	Method	Acc-2	F1 score	Acc-5	Acc-7	MAE	Corr
0.0	LNLN	84.14/82.96	82.49/81.38	52.42	50.94	0.566	0.769
	P-RMF	86.38/84.69	84.54/80.59	54.58	51.99	0.538	0.856
	ALMT	86.26/81.74	82.62/78.12	57.84	55.93	0.567	0.838
	Ours (HyDRA)	84.45/85.62	84.30/85.48	54.85	53.35	0.542	0.768
0.1	LNLN	83.48/81.99	81.57/80.08	51.56	50.31	0.582	0.744
	P-RMF	84.94/83.92	83.74/80.21	52.63	52.07	0.572	0.784
	ALMT	84.58/80.63	81.09/76.63	55.33	53.62	0.619	0.809
	Ours (HyDRA)	83.78/84.68	83.60/84.52	53.88	52.40	0.560	0.747
0.2	LNLN	81.68/80.98	79.98/79.14	50.44	49.07	0.609	0.708
	P-RMF	82.86/81.34	81.02/79.56	52.09	50.24	0.586	0.791
	ALMT	83.27/79.77	79.63/75.63	51.56	50.09	0.647	0.754
	Ours (HyDRA)	82.75/83.58	82.55/83.45	52.80	51.45	0.582	0.722
0.3	LNLN	80.23/79.81	78.33/78.18	48.78	47.69	0.642	0.657
	P-RMF	81.86/80.98	79.51/78.76	50.04	48.77	0.608	0.687
	ALMT	81.53/77.94	78.28/74.02	47.46	46.38	0.674	0.677
	Ours (HyDRA)	81.42/82.08	81.08/81.90	51.55	50.35	0.605	0.692
0.4	LNLN	78.67/78.97	77.08/77.83	47.21	46.34	0.663	0.621
	P-RMF	78.94/80.36	79.39/78.31	48.76	48.97	0.661	0.643
	ALMT	79.79/77.18	76.59/73.37	44.38	43.41	0.692	0.606
	Ours (HyDRA)	80.60/80.78	80.02/80.45	50.20	49.00	0.629	0.657
0.5	LNLN	76.41/77.22	75.24/76.87	45.87	45.21	0.702	0.553
	P-RMF	77.69/78.16	77.95/77.64	48.78	46.68	0.678	0.582
	ALMT	77.86/75.68	74.79/71.87	41.13	40.58	0.728	0.519
	Ours (HyDRA)	79.00/78.82	78.28/78.48	48.40	47.35	0.667	0.609
0.6	LNLN	74.04/75.89	73.26/75.72	44.31	43.74	0.723	0.492
	P-RMF	76.57/76.24	76.36/77.02	47.18	46.21	0.709	0.516
	ALMT	75.53/74.49	72.46/70.87	39.14	38.69	0.757	0.447
	Ours (HyDRA)	77.55/76.88	76.90/76.38	47.00	46.05	0.699	0.559
0.7	LNLN	71.66/74.08	71.63/75.01	43.04	42.48	0.756	0.426
	P-RMF	73.12/75.08	73.83/75.99	44.93	44.16	0.728	0.447
	ALMT	71.93/72.09	69.68/68.84	37.53	37.43	0.792	0.356
	Ours (HyDRA)	75.48/74.32	74.28/73.58	45.25	44.78	0.730	0.500
0.8	LNLN	68.57/72.27	69.99/74.34	41.24	40.99	0.779	0.342
	P-RMF	71.79/74.77	72.66/75.41	43.62	43.54	0.742	0.378
	ALMT	68.54/69.58	66.63/66.71	36.77	36.54	0.824	0.261
	Ours (HyDRA)	73.25/70.98	71.30/70.38	44.10	43.88	0.759	0.426
0.9	LNLN	64.87/70.81	68.78/74.97	40.63	40.26	0.818	0.228
	P-RMF	69.44/73.66	70.51/74.82	43.49	42.26	0.758	0.276
	ALMT	61.69/67.01	60.94/64.54	37.06	37.03	0.858	0.163
	Ours (HyDRA)	71.12/67.58	68.20/66.45	42.80	42.75	0.789	0.324
Avg.	LNLN	76.38/77.50	75.84/77.35	46.55	45.70	0.684	0.554
	P-RMF	78.36/78.92	77.95/77.83	48.61	47.49	0.658	0.596
	ALMT	77.10/75.61	74.27/72.06	44.82	43.97	0.716	0.543
	Ours (HyDRA)	78.94/78.53	78.05/78.11	49.08	48.14	0.656	0.600

Table 9: Complete comparison results on the MOSEI dataset.

Rate (r)	Method	Acc-2	F1 score	Acc-3	Acc-5	MAE	Corr
0.0	LNLN	76.94	73.97	62.44	37.48	0.484	0.503
	P-RMF	76.87	75.97	62.02	40.98	0.469	0.517
	ALMT	75.18	74.09	61.39	40.41	0.481	0.507
	Ours (HyDRA)	77.90	77.54	63.89	43.98	0.430	0.565
0.1	LNLN	76.09	72.81	61.36	36.74	0.493	0.472
	P-RMF	76.24	74.91	60.69	40.02	0.477	0.497
	ALMT	74.28	73.03	60.39	39.16	0.489	0.479
	Ours (HyDRA)	77.24	76.46	63.46	42.23	0.445	0.543
0.2	LNLN	75.34	71.57	59.59	35.42	0.507	0.438
	P-RMF	74.94	74.09	59.86	38.07	0.487	0.462
	ALMT	73.27	72.09	59.47	38.13	0.502	0.444
	Ours (HyDRA)	75.71	75.59	61.49	40.48	0.465	0.512
0.3	LNLN	74.09	70.52	58.86	33.92	0.517	0.398
	P-RMF	74.09	72.63	58.32	37.02	0.497	0.444
	ALMT	72.51	71.03	57.63	36.19	0.541	0.441
	Ours (HyDRA)	74.52	73.28	59.80	38.07	0.480	0.478
0.4	LNLN	74.84	71.33	59.38	34.76	0.511	0.416
	P-RMF	74.52	73.22	58.98	37.53	0.491	0.452
	ALMT	73.19	72.02	58.79	37.24	0.504	0.418
	Ours (HyDRA)	73.88	72.15	58.72	36.54	0.488	0.455
0.5	LNLN	72.91	69.73	57.06	32.41	0.526	0.342
	P-RMF	73.24	71.41	56.78	35.31	0.508	0.409
	ALMT	71.66	70.39	56.79	34.64	0.461	0.328
	Ours (HyDRA)	72.95	71.42	57.55	34.79	0.496	0.433
0.6	LNLN	71.48	68.11	55.62	30.93	0.543	0.344
	P-RMF	71.94	70.12	54.83	33.19	0.518	0.378
	ALMT	70.31	68.78	55.14	32.47	0.531	0.321
	Ours (HyDRA)	72.15	70.31	56.26	32.60	0.509	0.409
0.7	LNLN	70.16	66.53	53.02	28.79	0.557	0.256
	P-RMF	70.21	68.27	52.96	30.68	0.537	0.329
	ALMT	69.07	67.22	52.97	29.71	0.551	0.294
	Ours (HyDRA)	71.29	69.55	55.10	31.51	0.524	0.359
0.8	LNLN	68.56	64.41	50.83	26.61	0.579	0.212
	P-RMF	68.44	66.44	50.62	28.38	0.551	0.286
	ALMT	67.74	65.72	50.92	27.92	0.569	0.222
	Ours (HyDRA)	70.35	68.20	53.80	30.20	0.550	0.315
0.9	LNLN	68.29	64.22	50.64	26.04	0.583	0.209
	P-RMF	68.41	65.94	50.54	27.92	0.556	0.276
	ALMT	67.29	65.03	50.61	27.33	0.601	0.196
	Ours (HyDRA)	68.81	65.10	55.03	24.71	0.575	0.234
Avg.	LNLN	72.87	69.32	56.88	32.31	0.530	0.359
	P-RMF	72.89	71.30	56.56	34.91	0.509	0.405
	ALMT	71.45	69.94	56.41	34.32	0.523	0.365
	Ours (HyDRA)	73.48	71.96	58.51	35.51	0.496	0.430

Table 10: Complete comparison results on the SIMS dataset.

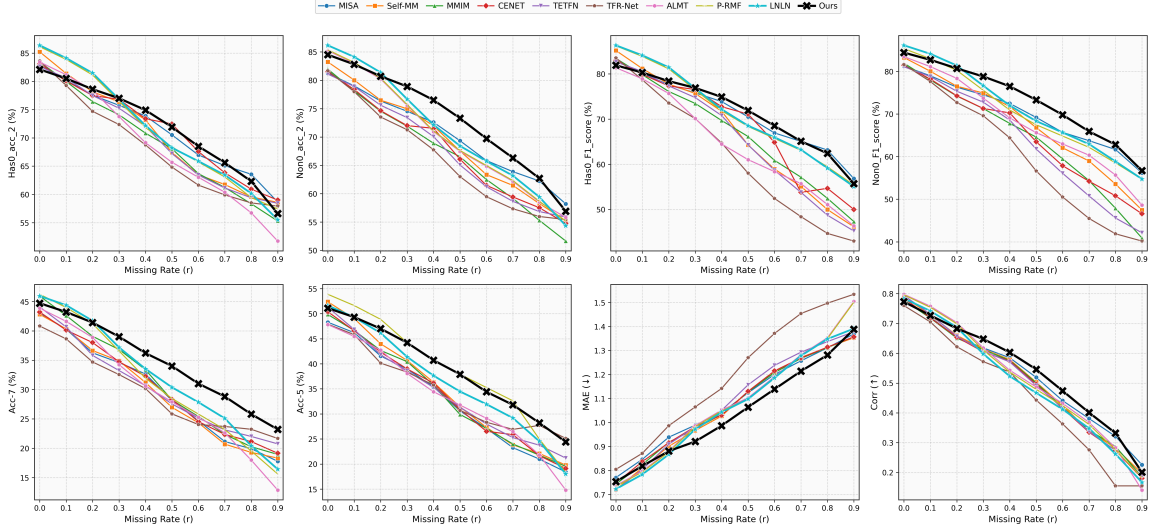


Figure 7: Performance comparison across all metrics on the MOSI dataset under varying missing rates.

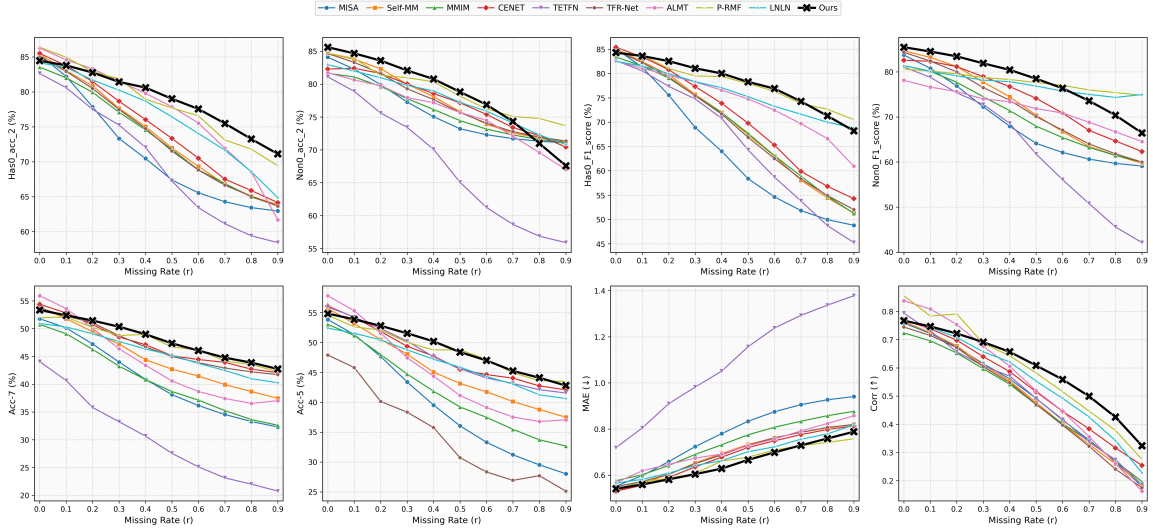


Figure 8: Performance comparison across all metrics on the MOSEI dataset under varying missing rates.

Method	MAE↓	Corr↑
HyDRA	1.044±0.011	0.539±0.008
LNLN	1.069±0.014	0.499±0.010
P-RMF	1.085±0.016	0.512±0.011

Table 11: Run-to-run stability on MOSI under the independent five-seed protocol {42, 123, 2024, 7, 31}. Results are reported as mean $\pm$ standard deviation.

the between-method gap on every row, matching the tight run-to-run stability reported in Table 11.

F Hyperparameter and Training Sensitivity

We analyze the sensitivity of HyDRA to key hyperparameters on the MOSI dataset, computing average test results across missing rates 0–0.9. These

experiments are included to check whether the proposed gains rely on fragile hyperparameter choices. We group the checks into two compact summaries: Table 13 covers geometry and distillation parameters, Table 14 covers trainable curvature variants, and Table 15 covers auxiliary losses and training schedules.

F.1 Geometry and Distillation Parameters

We first vary the Poincaré curvature c , the maximum distillation weight $\beta_{\max}$, and trainable curvature parameterisations. Recent hyperbolic learning work (Ganea et al., 2018; Gulcehre et al., 2019; Shimizu et al., 2021) reports that making c trainable can improve some graph and language tasks, so we include learnable scalar, per-layer, and per-sample curvature variants while keeping every other com-

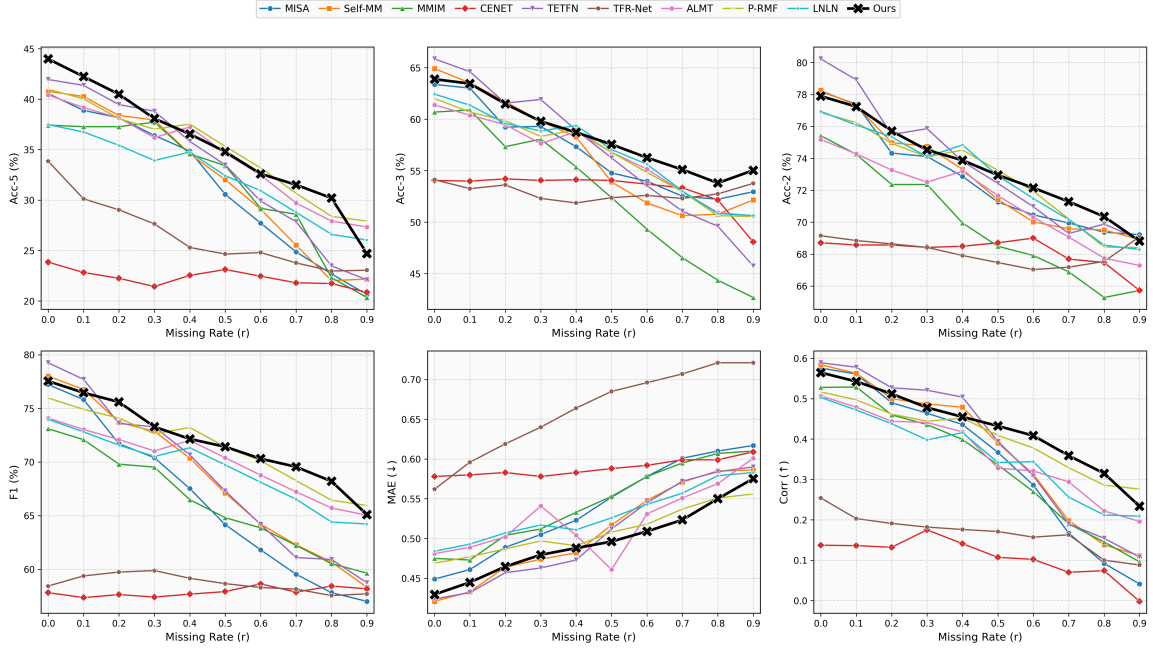


Figure 9: Performance comparison across all metrics on the SIMS dataset under varying missing rates.

Dataset	Metric	Δ	p -value	95% CI
MOSI	Acc-2 (Has0)	+0.46	0.029	[+0.09, +0.78]
	Acc-2 (Non0)	+2.80	< 0.001	[+2.41, +3.18]
	Acc-7	+3.48	0.002	[+2.25, +4.67]
	MAE $\downarrow$	-0.040	0.004	[-0.054, -0.025]
MOSEI	Acc-2 (Has0)	+0.58	0.028	[+0.12, +1.02]
	Acc-5	+0.47	0.023	[+0.11, +0.81]
	Acc-7	+0.65	0.007	[+0.27, +1.01]
	F1 (Non0)	+0.28	0.034	[+0.04, +0.52]
SIMS	Acc-2	+0.59	0.031	[+0.10, +1.05]
	Acc-3	+1.95	0.003	[+1.26, +2.62]
	F1	+0.66	0.018	[+0.20, +1.10]

Table 12: Paired significance tests between HyDRA and P-RMF under the independent five-seed protocol. Δ is the mean improvement (absolute, in percentage points except MAE), p -values are from two-sided paired t -tests, and 95% CIs are paired-bootstrap intervals with $B=10,000$ resamples. All 11 metrics yield $p < 0.05$ and confidence intervals excluding zero.

ponent fixed.

Findings. Table 13 shows that $c = 1.0$ and $c = 2.0$ are close, but $c = 1.0$ is better on MAE, Corr, and Acc-7 in the fixed-curvature setting, so we keep the standard curvature for stability. The same table shows that $\beta_{\max} = 1.0$ gives the best balance between distillation and task supervision. Trainable curvature brings only negligible changes: the learnable scalar converges to $c^* = 1.03 \pm 0.12$ on MOSI and 0.97 ± 0.09 on MOSEI, while the per-layer variant converges near $(1.06, 0.94)$ and slightly improves a few metrics by at most 0.02 Acc-7, 0.002 MAE, and 0.001 Corr. We therefore keep fixed $c = 1.0$ because it is simpler and com-

parably strong. The per-sample variant degrades performance, suggesting that coupling curvature directly to completeness is redundant with DAHP’s radial positioning.

F.2 Objective Weights and Training Schedule

Because HyDRA combines task prediction, completeness scoring, hyperbolic correction, and feature reconstruction, we further test whether the auxiliary loss weights and the progressive correction schedule dominate the final performance.

Table 15 shows that performance remains stable around the selected auxiliary weights. Very small λ_1 weakens completeness scoring, while overly large λ_2 gives too much emphasis to reconstruction.

Hyperparameter	Setting	Acc-2	F1 score	Acc-5	Acc-7	MAE↓	Corr↑
c	0.5	71.23/72.58	71.04/72.47	37.89	33.65	1.068	0.521
	1.0	71.80 /73.21	71.61 /73.16	38.90	34.73	1.044	0.539
	2.0	71.56/ 73.34	71.38/ 73.29	39.12	34.45	1.052	0.532
	5.0	70.67/72.14	70.48/72.03	37.21	33.08	1.089	0.508
$\beta_{\max}$	0.5	71.12/72.67	70.93/72.56	37.78	33.89	1.073	0.518
	1.0	71.80 /73.21	71.61 /73.16	38.90	34.73	1.044	0.539
	2.0	71.67/73.08	71.49/72.98	39.05	34.38	1.058	0.531
	5.0	70.45/71.89	70.26/71.78	36.87	32.76	1.098	0.497

Table 13: Sensitivity to fixed curvature and maximum distillation weight on MOSI. Results are averaged over missing rates 0–0.9. Best results are in **bold**.

Variant	MOSI				MOSEI			
	Acc-7	MAE↓	Corr↑	#params [†]	Acc-7	MAE↓	Corr↑	#params [†]
Fixed $c = 1.0$ (ours)	34.73	1.044	0.539	0	48.14	0.656	0.600	0
Learnable scalar c	34.69	1.046	0.538	1	48.09	0.657	0.599	1
Per-layer c_ℓ	34.71	1.042	0.540	2	48.16	0.655	0.601	2
Per-sample $c(\hat{w})$	34.58	1.049	0.535	2,113	47.92	0.661	0.596	2,113

Table 14: Trainable curvature variants averaged over missing rates 0–0.9 and three seeds. [†] denotes extra parameters over fixed $c = 1.0$. Best results are in **bold**.

Hyperparameter	Setting	MAE↓	Corr↑
λ_1	0.1	1.076	0.516
	0.5	1.057	0.528
	0.9	1.044	0.539
	1.5	1.053	0.532
λ_2	0.01	1.063	0.526
	0.1	1.044	0.539
	0.5	1.056	0.530
	1.0	1.069	0.521
T_{warmup}	0	1.103	0.491
	5	1.058	0.533
	10	1.044	0.539
	20	1.053	0.536
Schedule	Linear	1.044	0.539
	Cosine	1.051	0.533
	Exponential	1.061	0.525

Table 15: Sensitivity of auxiliary loss weights and training schedule on MOSI. Results are averaged over missing rates 0–0.9. λ_1 controls completeness supervision, λ_2 controls reconstruction, T_{warmup} delays hyperbolic correction, and schedule denotes the progressive distillation schedule.

For the schedule, applying hyperbolic correction from the beginning is less stable, an excessively long warmup delays teacher guidance, and the linear schedule is the best among the tested options.

G Experimental Protocol

G.1 Dataset Details

We evaluate our proposed method on three widely-used benchmark datasets for multimodal sentiment analysis. These datasets encompass both English and Chinese languages, varying scales, and diverse topic coverage, providing a comprehensive testbed for assessing the effectiveness and generalizability of our approach.

- **MOSI**: The CMU-MOSI dataset contains speakers’ opinion expressions on topics such as movies from 93 YouTube videos, comprising 2,199 video clips (Zadeh et al., 2016). The sentiment intensity of each clip is annotated in the range of -3 (strongly negative) to +3 (strongly positive).
- **MOSEI**: The CMU-MOSEI dataset is an extended version of MOSI, covering 250 different topics with 22,856 annotated video clips (Zadeh et al., 2018). The sentiment annotation method is the same as MOSI.
- **SIMS**: The SIMS dataset is a Chinese multimodal sentiment analysis dataset containing 2,281 video clips (Yu et al., 2020). Each sample has one multimodal sentiment label and three unimodal sentiment labels, with sentiment scores ranging from -1 to +1.

Dataset	Speaker	Video Clip	Train	Valid	Test	Language
MOSI	93	2199	1284	229	686	English
MOSEI	1000	22856	16326	1871	4659	English
SIMS	474	2281	1368	456	457	Chinese

Table 16: Dataset statistics.

G.2 Baseline Details

(1) **Complete modality methods** assume all modalities are fully available during both training and inference:

- **MISA** (Hazarika et al., 2020) projects each modality into invariant and specific subspaces in Euclidean space, disentangling shared and unique sentiment factors through linear decomposition;

- **Self-MM** (Yu et al., 2021) generates unimodal supervision labels through a self-supervised multi-task scheme, enabling each modality branch to learn discriminative sentiment representations independently *without cross-modal information exchange*;
- **MMIM** (Han et al., 2021) maximises mutual information hierarchically—first between unimodal pairs and then between the fused output and each individual modality—under a *Euclidean statistical framework* to strengthen inter-modal coherence;
- **CENET** (Wang et al., 2022) enriches the dominant linguistic stream with auxiliary acoustic and visual signals through pairwise bimodal interaction layers with *fixed interaction weights*;
- **TETFN** (Wang et al., 2023) performs unidirectional text-to-nonverbal enhancement in Euclidean space by attending acoustic and visual branches to textual sentiment cues;
- **ALMT** (Zhang et al., 2023) dynamically modulates non-verbal contributions conditioned on linguistic context through *modality-level soft gating* in Euclidean space to filter out sentiment-irrelevant and conflicting information from auxiliary modalities.

(2) **Incomplete-modality methods** are specifically designed to handle incomplete multimodal data:

- **TFR-Net** (Yuan et al., 2021) employs transformers to impute absent modality tokens in temporally non-aligned sequences through *direct feature-level generation* in Euclidean space;
- **LNLN** (Zhang et al., 2024a) prioritises the linguistic signal as a noise-resistant anchor and applies dominant-modality calibration at a *uniform strength* to attenuate the impact of absent or corrupted non-verbal inputs;
- **P-RMF** (Zhu et al., 2025) learns canonical joint embeddings from partially observed inputs in Euclidean space, scores each modality’s reliability to re-weight its fusion contribution, and performs *pairwise cross-modal interaction* for robustness under missingness;
- **DAR** (Yang et al., 2025) factorises each modality into shared and private subspaces via mutual-

information constraints and recombines the available factors to fill in absent components by applying *uniform constraints across all samples*. Since the official DAR implementation is not publicly available and its published results do not cover the full missing-rate grid under our unified protocol, re-implementing DAR from the paper alone would introduce uncontrolled differences in preprocessing, masking, and training details. We therefore report DAR only in the partial comparisons where comparable published results are available.

Reproduction protocol. We reproduced ALMT, LNLN, and P-RMF under identical dataset splits and preprocessing pipelines; results for the remaining baselines are cited from Zhang et al. (2024a), which follows the same experimental protocol.

G.3 Missing Protocol

To simulate real-world missingness caused by sensor failures or ASR errors, we follow the protocol of Zhang et al. (2024a) and randomly erase a fraction $r \in [0, 0.9]$ of information in each modality for every sample. The per-rate evaluation grid is $r \in \{0.0, 0.1, \dots, 0.9\}$. Zero padding is applied to the erased portions of the visual and audio modalities, while the text modality is filled with the [UNK] token. Formally, the incomplete input is obtained by:

$$\tilde{U}^m = M^m \odot U^m \quad (19)$$

where $M^m \in \{0, 1\}^{T_m \times d_m}$ is a binary mask and $\odot$ denotes element-wise multiplication.

For MOSI and MOSEI, binary metrics use the Has0/Non0 convention: Has0 includes neutral samples in the non-negative class, whereas Non0 excludes neutral samples before computing Acc-2 and F1.

G.4 Training Details and Statistical Protocol

The text modality is encoded using BERT-base (Devlin et al., 2019), visual features are extracted via OpenFace 2.0 (Baltrušaitis et al., 2018), and audio features are obtained using Librosa (McFee et al., 2015). We use the AdamW optimizer with learning rate $\eta = 10^{-4}$ and weight decay 10^{-4} . To stabilize hyperbolic space optimization, gradient norm clipping is applied with threshold 1.0.

Unless otherwise stated, all newly generated results for HyDRA, reproduced baselines, ablations, and sensitivity variants use random seeds {1111, 1112, 1113} and are averaged across seeds

and missing rates $r \in \{0.0, 0.1, \dots, 0.9\}$. This three-seed protocol follows Zhang et al. (2024a) and keeps the main comparison, ablations, and sensitivity analyses under the same data splits, preprocessing pipeline, and evaluation grid. Results for externally cited baselines follow the same evaluation protocol when available.

To separately assess statistical reliability rather than reproduction fidelity, we additionally re-run HyDRA and the closest competitor with an independent set of five random seeds {42, 123, 2024, 7, 31}, disjoint from the canonical three seeds. We report paired t -tests, paired-bootstrap 95% confidence intervals, and run-to-run standard deviations in Appendix E.1–E.1. The two tiers answer complementary questions: the three-seed tier keeps the main comparison aligned with prior work, while the five-seed tier supplies multi-seed significance evidence without re-tuning or cherry-picking seeds.

G.5 Hyperparameter Settings

Table 17 summarizes the hyperparameter settings for all datasets.

Hyperparameter	MOSI	MOSEI	SIMS
Language input dimension d_l	768	768	768
Audio input dimension d_a	5	74	33
Visual input dimension d_v	20	35	709
Hyperbolic space dimension d'	128	128	128
Poincaré ball curvature c	1.0	1.0	1.0
Batch size	64	64	64
Epochs	200	200	200
Learning rate η	1e-4	1e-4	1e-4
$\lambda_1, \beta_{\max}, \lambda_2$	0.9, 1.0, 0.1	0.9, 1.0, 0.1	0.9, 1.0, 0.1
Warmup epochs T_{warmup}	10	10	10
EMA momentum μ	0.999	0.999	0.999
Slack margin τ	0.1	0.1	0.1
Safety margin ϵ	1e-5	1e-5	1e-5

Table 17: Hyperparameter settings for all datasets.

H Hyperbolic Formulation and Mathematical Motivation

This appendix provides detailed formulations of hyperbolic geometric operations used in the main text, including the exponential map $\exp_0$, logarithmic map $\log_0$, Möbius addition $\oplus_c$, hyperbolic distance $d_{\mathbb{H}}$, and boundary projection Π_c .

H.1 Poincaré Ball Model

The Poincaré ball model $\mathbb{D}_c^n = \{z \in \mathbb{R}^n : c\|z\|^2 < 1\}$ is a hyperbolic space model with curvature $-c$ ($c > 0$), where n denotes the embedding dimension (corresponding to d' in the main text). Compared

to Euclidean space, hyperbolic space exhibits exponential volume growth near the boundary, allowing it to separate samples of different quality levels with greater resolution in low dimensions (Nickel and Kiela, 2017). This property makes hyperbolic space suitable for modeling the representation drift that grows much faster at high missing rates than at low missing rates (see Appendix A.1).

H.2 Exponential and Logarithmic Maps

The exponential map $\exp_0 : T_0\mathbb{D}_c^n \rightarrow \mathbb{D}_c^n$ projects tangent vectors $v \in T_0\mathbb{D}_c^n \cong \mathbb{R}^n$ from Euclidean space to hyperbolic space:

$$\exp_0(v) = \tanh(\sqrt{c}\|v\|) \frac{v}{\sqrt{c}\|v\|} \quad (20)$$

This operation is used in Eq. (5) of the main text to map the fused Euclidean representation G into a content-dependent hyperbolic proposal before completeness-driven radial positioning.

The logarithmic map $\log_0 : \mathbb{D}_c^n \rightarrow T_0\mathbb{D}_c^n$ performs the inverse operation, mapping points z from hyperbolic space back to the tangent space:

$$\log_0(z) = \frac{\operatorname{artanh}(\sqrt{c}\|z\|)}{\sqrt{c}\|z\|} z \quad (21)$$

where $\operatorname{artanh}(\cdot)$ denotes the inverse hyperbolic tangent function. For numerical stability, when $\|z\| \rightarrow 0$, we have $\log_0(z) \rightarrow z$.

H.3 Möbius Addition

Möbius addition $\oplus_c : \mathbb{D}_c^n \times \mathbb{D}_c^n \rightarrow \mathbb{D}_c^n$ defines vector combination in hyperbolic space, generalizing Euclidean addition to account for curved geometry:

$$x \oplus_c y = \frac{(1 + 2c\langle x, y \rangle + c\|y\|^2)x}{1 + 2c\langle x, y \rangle + c^2\|x\|^2\|y\|^2} + \frac{(1 - c\|x\|^2)y}{1 + 2c\langle x, y \rangle + c^2\|x\|^2\|y\|^2} \quad (22)$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product. This operation satisfies key properties: (1) $x \oplus_c \mathbf{0} = x$ (identity), and (2) $x \oplus_c (-x) = \mathbf{0}$ (inverse). As $c \rightarrow 0$, Möbius addition reduces to standard Euclidean addition.

H.4 Hyperbolic Distance

Based on Möbius addition, the hyperbolic distance between two points $x, y \in \mathbb{D}_c^n$ is defined as:

$$d_{\mathbb{H}}(x, y) = \frac{2}{\sqrt{c}} \operatorname{artanh}(\sqrt{c}\| -x \oplus_c y \|) \quad (23)$$

This distance metric is used in the hyperbolic distillation loss $\mathcal{L}_{\text{hyp}}$ (Eq. (11)). A key property is that hyperbolic distance exhibits significant amplification near the boundary of the Poincaré ball: for points with $\|z\| \rightarrow 1/\sqrt{c}$, even small Euclidean displacements result in large hyperbolic distances. This boundary sensitivity provides stronger corrective gradients for severely missing or low-completeness samples.

H.5 Boundary Projection

The projection operator $\Pi_c : \mathbb{R}^n \rightarrow \mathbb{D}_c^n$ ensures numerical stability by constraining points within the Poincaré ball:

$$\Pi_c(z) = \begin{cases} z & \text{if } \sqrt{c}\|z\| < 1 - \epsilon \\ \frac{(1 - \epsilon)}{\sqrt{c}\|z\|} z & \text{otherwise} \end{cases} \quad (24)$$

where $\epsilon = 10^{-5}$ is a small constant preventing numerical overflow when computing hyperbolic operations near the boundary. This projection is applied after completeness-driven radial positioning in Eq. (9) to guarantee that all representations remain within the valid hyperbolic space.

H.6 Riemannian Metric and Conformal Factor

The Poincaré ball is a conformal model of hyperbolic space, meaning angles are preserved but distances are scaled by a position-dependent conformal factor. The Riemannian metric tensor at point $z \in \mathbb{D}_c^n$ is given by:

$$g_z = \lambda_z^2 I_n, \quad \text{where } \lambda_z = \frac{2}{1 - c\|z\|^2} \quad (25)$$

Here, λ_z is the conformal factor that governs the local geometry. As $\|z\| \rightarrow 1/\sqrt{c}$ (approaching the boundary), $\lambda_z \rightarrow \infty$, causing distances and gradients to be strongly amplified. This boundary amplification effect is the key geometric property exploited by our method: low-completeness samples placed near the boundary experience substantially stronger gradient signals during optimization.

H.7 Summary of Notation

Table 18 summarizes the hyperbolic geometric notation used throughout this paper, and Table 19 lists the HyDRA-specific symbols introduced in § 3.2–3.4 together with their tensor shapes and admissible ranges. The two tables are grouped here so that the main text can refer to any symbol without interrupting the flow of § 3.

Table 18: Summary of hyperbolic geometric notation.

Symbol	Description
$\mathbb{D}_c^n$	Poincaré ball (dim n , curvature $-c$)
c	Curvature parameter ($c > 0$)
$\exp_0$	Exponential map at origin
$\log_0$	Logarithmic map at origin
$\oplus_c$	Möbius addition
$d_{\mathbb{H}}$	Hyperbolic distance
Π_c	Boundary projection operator
λ_z	Conformal factor at point z

Table 19: HyDRA-specific symbols introduced in § 3.2–3.4, grouped by module.

Symbol	Meaning	Shape / Range
<i>Inputs and CMER (§ 3.2)</i>		
$U^m, \tilde{U}^m$	complete / incomplete input	$T_m \times d_m$
M^m	binary missingness mask	$\{0, 1\}^{T_m \times d_m}$
r_m	modality missing rate	$[0, 1]$
H^l, H^a, H^v	encoded L / A / V feats.	$n \times d$
$\mathbf{P}$	learnable language proxy	$n \times d$
H_{proxy}	proxy-retrieved A/V evidence	$n \times d$
g	scalar gating coefficient	$[0, 1]$
$\hat{H}^l$	compensated language repr.	$n \times d$
<i>DAHP (§ 3.3)</i>		
$\hat{w}$	estimated completeness score	$[0, 1]$
w	supervision target $1 - r_l$	$[0, 1]$
$\mathbf{h}_{(\ell)}$	semantic token at layer ℓ	$n \times d$
G	fused semantic vector	d
$\tilde{z}$	content hyperbolic proposal	$\mathbb{D}_c^{d'}$
u	unit semantic direction	$\mathbb{R}^{d'}$
$\rho(\hat{w})$	damage-controlled radius (after Π_c)	$[0, \rho_{\max}]$
$\rho_{\max}$	boundary $(1 - \epsilon) / \sqrt{c}$	$\mathbb{R}_{>0}$
z	damage-aware position	$\mathbb{D}_c^{d'}$
<i>SAHC (§ 3.4)</i>		
z_S, z_T	student / EMA teacher	$\mathbb{D}_c^{d'}$
θ_S, θ_T	student / teacher params.	–
μ	EMA momentum	$[0, 1]$
τ	relaxation margin	$\mathbb{R}_{\geq 0}$
$\hat{q}$	damage score $1 - \hat{w}$	$[0, 1]$
<i>Learning objective (§ 3.5)</i>		
λ_1, λ_2	loss balancing coeffs.	$\mathbb{R}_{\geq 0}$
$\beta(t)$	progressive distill. weight	$[0, \beta_{\max}]$
T_{warmup}	warmup epochs for $\beta(t)$	$\mathbb{N}$

H.8 Intuitive Reading of DAHP Radial Equations

Equations (6)–(9) in the main text couple the completeness score $\hat{w}$ with the hyperbolic coordinate z through three steps: direction extraction, radius scheduling, and boundary projection. We restate them here with a plain-language interpretation, so that a reader can parse § 3.3 without re-deriving the Poincaré algebra.

- **Direction carries semantics.** $u = \tilde{z} / (\|\tilde{z}\|_2 + \epsilon)$ keeps the sentiment-bearing direction of the fused proposal unchanged. Two samples with the same u but different $\hat{w}$ will be read the same way by the predictor, so the downstream task is not perturbed by radial re-positioning.
- **Radius carries damage.** $\rho(\hat{w}) = \hat{w}\|\tilde{z}\|_2 + (1 - \hat{w})\rho_{\max}$ linearly interpolates between the content-native radius $\|\tilde{z}\|_2$ (when $\hat{w} = 1$, i.e. intact input) and the boundary radius $\rho_{\max}$ (when $\hat{w} = 0$, i.e. maximally damaged input). The mapping is monotone decreasing in completeness, so more severely damaged samples are pushed farther toward the boundary.
- **Projection keeps the geometry valid.** $z = \Pi_c(\rho(\hat{w})u)$ applies the boundary projection of Eq. (24), ensuring $\sqrt{c}\|z\| < 1$ even under numerical noise. Inside the ball, z shares the direction of $\tilde{z}$ and the magnitude of $\rho(\hat{w})$.
- **Why the boundary matters.** Because the conformal factor $\lambda_z = 2/(1 - c\|z\|^2)$ in Eq. (25) diverges as $\|z\| \rightarrow 1/\sqrt{c}$, a small Euclidean displacement near the boundary becomes a large hyperbolic displacement. Coupling this with the distillation weight $1 - \hat{w}$ in Eq. (11) gives HyDRA its central property: *samples with high completeness stay near the semantic core and receive gentle correction; samples with low completeness are radially expanded to the boundary, where the geometry itself amplifies the correction signal.*

This paragraph is the intuition companion of the formal derivation in Appendix H.9: the equations here describe *what* the model does, whereas Eq. (32) explains *why* the resulting correction coefficient is monotone in damage.

H.9 Mathematical Motivation of HyDRA

This appendix gives the mathematical motivation behind the two geometric designs used in HyDRA: completeness-controlled radial positioning

and severity-adaptive hyperbolic correction. The goal is not to prove global optimality, but to show that the proposed mechanism gives a monotone and geometry-aware correction signal as sample damage increases. We therefore analyze a local slice of the model where the semantic direction is fixed, the boundary projection is inactive, and the hinge loss is active when studying correction strength.

H.9.1 Radial Severity from Completeness

Let $\hat{q} = 1 - \hat{w}$ denote the estimated damage score, where $\hat{w} \in [0, 1]$ is the predicted completeness. For a semantic proposal $\tilde{z}$ obtained from text-guided fusion, define $r_0 = \|\tilde{z}\|_2$. In the nonzero case, its semantic direction is $u = \tilde{z}/\|\tilde{z}\|_2$; the implementation in Eq. (9) adds ϵ only to avoid division by zero. DAHP assigns the final radius as

$$\rho(\hat{q}) = (1 - \hat{q})r_0 + \hat{q}\rho_{\max} = r_0 + \hat{q}(\rho_{\max} - r_0), \quad (26)$$

where $\rho_{\max} = (1 - \epsilon)/\sqrt{c}$. Since Π_c clips saturated points to the valid radius, we consider the non-saturated regime $0 \leq r_0 < \rho_{\max}$; if r_0 is already clipped, the sample has reached the maximum radial severity. In the non-saturated regime,

$$\frac{\partial \rho(\hat{q})}{\partial \hat{q}} = \rho_{\max} - r_0 > 0. \quad (27)$$

Therefore, lower completeness always pushes the sample farther from the origin while preserving the direction u . This separates two roles: u carries sentiment semantics for prediction, while $\rho(\hat{q})$ encodes how severely the sample has been damaged.

The corresponding hyperbolic radial severity is the distance from the origin:

$$s(\hat{q}) = d_{\mathbb{H}}(0, z) = \frac{2}{\sqrt{c}} \operatorname{artanh}(\sqrt{c}\rho(\hat{q})). \quad (28)$$

Differentiating Eq. (28) gives

$$\frac{\partial s(\hat{q})}{\partial \hat{q}} = \frac{2(\rho_{\max} - r_0)}{1 - c\rho(\hat{q})^2} = \lambda_{\rho(\hat{q})}(\rho_{\max} - r_0) > 0, \quad (29)$$

where $\lambda_{\rho} = 2/(1 - c\rho^2)$ is the conformal factor in Eq. (25). As $\rho(\hat{q})$ approaches the boundary, the denominator becomes smaller, so the same increase in damage produces a larger increase in hyperbolic severity. This explains why DAHP maps severely damaged samples toward the boundary rather than only assigning them a larger scalar weight.

H.9.2 Effective Strength of Hyperbolic Correction

SAHC uses the loss

$$\mathcal{L}_{\text{hyp}} = \hat{q} \max(0, d_{\mathbb{H}}(z_S, z_T) - \tau), \quad (30)$$

where z_S is the student representation from incomplete inputs and z_T is the complete-modality teacher anchor. For local comparative statics, we fix z_T and the semantic direction of z_S , so the radius of z_S follows $\rho(\hat{q})$. When the hinge is active, a local displacement $\Delta = z_S - z_T$ satisfies the first-order approximation

$$d_{\mathbb{H}}(z_S, z_T) \approx \lambda_{z_S} \|\Delta\|_2. \quad (31)$$

Combining Eq. (31) with the radial positioning in Eq. (26), the local effective correction coefficient can be written as

$$\Gamma(\hat{q}) = \hat{q} \lambda_{\rho(\hat{q})} = \hat{q} \frac{2}{1 - c\rho(\hat{q})^2}. \quad (32)$$

Its derivative is

$$\frac{\partial \Gamma(\hat{q})}{\partial \hat{q}} = \lambda_{\rho(\hat{q})} + \hat{q} \frac{4c\rho(\hat{q})(\rho_{\max} - r_0)}{(1 - c\rho(\hat{q})^2)^2} > 0. \quad (33)$$

Thus, the correction strength grows with the estimated damage score in two coupled ways: the explicit factor $\hat{q}$ increases the distillation weight, and the radial movement increases the conformal factor $\lambda_{\rho(\hat{q})}$. A scalar-weighted Euclidean distillation loss can provide the first factor, but not the second position-dependent amplification, making it less suited to the drift that varies across missing rates and samples.

H.9.3 Scope of the Derivation

The derivation above supports the mechanism used by HyDRA: completeness controls the radial coordinate, the radial coordinate controls the hyperbolic metric, and the metric determines the local correction strength. The local approximation in Eq. (31) is used only to interpret optimization dynamics; the actual model is trained with the exact hyperbolic distance in Eq. (11). Therefore, the appendix should be read as a mathematical motivation for the adaptive correction design rather than as a claim of global optimization guarantees.

I Efficiency Analysis

To evaluate the computational efficiency of HyDRA, we compare it with the LNLN baseline on

Metric	LNLN	HyDRA (Ours)
Parameters (M) ↓	115.97	113.34
Train Time (s/epoch) ↓	4.22	5.77
Inference Latency (ms) ↓	71.02	85.87

Table 20: Efficiency comparison between LNLN and HyDRA on the MOSI dataset. ↓ indicates lower is better.

the MOSI dataset in terms of parameter count, training time per epoch, and inference latency. The results are summarized in Table 20.

As presented in Table 20, HyDRA uses fewer parameters than LNLN, indicating that its performance gains do not come from simply scaling up the model size. This compactness is consistent with the representational capacity of hyperbolic geometry, which can encode a wide range of damage levels efficiently.

Regarding computational speed, HyDRA incurs moderate overhead in training time and inference latency, mainly due to the additional completeness-aware fusion mechanism and Riemannian operations. The latency remains below 100 ms in our setting, so the added cost is manageable relative to the robustness gains under incomplete-modality conditions.

J Case Study

To intuitively demonstrate the effectiveness of HyDRA, we present qualitative examples in Figure 10, showing both successful and failed predictions under various incomplete-modality conditions. In successful cases, our method accurately captures sentiment even with severe missingness, demonstrating that cross-modal compensation and damage-aware hyperbolic correction effectively handle incomplete data.

K AI Assistants Disclosure

We used AI assistants to assist with polishing the English writing and grammar checking of the manuscript. The AI tools were not used to generate core research ideas, experimental designs, or results. All scientific content, analysis, and conclusions were produced entirely by the authors.

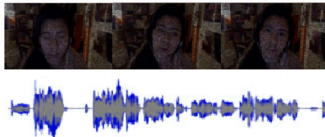
Input	MY 8 YEAR OLD WAS TERRIFIED THROUGH PART OF THIS BECAUSE THE ORIGINAL VERSION IS TERRIFYING 	SIGH I LOVE WHEN THEY SHOW ALL THE THE WAR SCENES ALL THE THE TIME THEY SPENT IN THE MILITARY AND SHIT 	AND, UM, ILIKED, YOU KNOW, CATHERINE HARDWICKE IS, LIKE, YOU KNOW, KIND OF RIDICULOUS 
	ALMT	Positive×	Positive×
LNLN	Negative✓	Negative✓	Positive×
H ² -CAD	Negative✓	Negative✓	Negative✓

Figure 10: Illustrative case study.