LFUF: CONTEXT-GUIDED LÉVY FLOW RECOVERY AND UNCERTAINTY-AWARE EVIDENCE FUSION FOR INCOMPLETE MULTIMODAL EMOTION RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal Emotion Recognition in Conversations (MERC) aims to infer utterance-level emotions from complementary linguistic, acoustic, and visual cues. In real-world scenarios, however, one or more modalities may be unavailable, leading to substantial information loss. Moreover, multimodal observations are inherently heterogeneous in informativeness and reliability, and different modalities associated with the same utterance may provide inconsistent or even conflicting emotional evidence. Consequently, directly recovering and fusing multimodal representations can propagate unreliable information and degrade prediction robustness. To address these challenges, we propose **Lévy Flow and Uncertainty-Aware Fusion (LFUF)**, a unified framework for context-guided missing-modality recovery and reliability-aware evidence fusion. LFUF first exploits speaker identities and conversational dependencies to construct contextual priors, which, together with the available modalities, condition a Lévy-stable flow-matching process for recovering missing representations. The observed and recovered representations are then converted into complementary evidence sources rather than being fused indiscriminately. To account for source reliability, LFUF quantifies inter-source consistency using the Jousselme distance and adaptively discounts unreliable evidence by reallocating part of its belief mass to the uncertainty set. Finally, local conflicts are explicitly resolved through PCR6-based evidence redistribution, yielding a conflict-aware fused belief for emotion prediction. Extensive experiments under diverse missing-modality settings demonstrate that LFUF consistently improves emotion recognition performance and exhibits strong robustness to incomplete and conflicting multimodal observations.

1 INTRODUCTION

Multimodal Emotion Recognition in Conversations (MERC) aims to identify the emotion of each utterance by jointly exploiting linguistic, acoustic, and visual cues Wu et al. (2025); Hu et al. (2021). Compared with unimodal analysis, multimodal perception captures complementary semantic, prosodic, and behavioral information, enabling a more comprehensive understanding of human affect. MERC has broad applications in empathetic dialogue systems, intelligent customer service, educational assistance, social media analysis, and affect-aware human–computer interaction. Improving its accuracy and robustness is therefore important for both affective computing research and real-world emotion-aware systems.

Most existing MERC methods assume that all modalities are fully observed. In practice, however, speech recognition errors, sensor failures, visual occlusion, privacy constraints, and unstable transmission can lead to partially missing modalities. Moreover, modalities differ in informativeness and reliability and may provide conflicting emotional evidence for the same utterance. Existing studies mainly address these issues through robust representation learning, missing-modality recovery, relational modeling, and adaptive multimodal fusion (Zhao et al., 2021; Ma et al., 2022; Lin & Hu, 2023; Guo et al., 2024). While these methods alleviate the impact of missing information, recovered representations may still contain substantial uncertainty or distortion. Conventional fusion schemes also often lack explicit mechanisms for identifying unreliable evidence and resolving inter-source conflicts, allowing recovery errors to propagate into the final prediction.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Recent methods have made substantial progress in incomplete multimodal emotion recognition. MMIN (Zhao et al., 2021) uses cross-modal imagination and cycle-consistent reconstruction to infer missing representations from available modalities. GCNet (Lian et al., 2023) models conversational relations with graph structures while jointly performing feature reconstruction and emotion classification. IMDer (Wang et al., 2023) employs conditional score-based diffusion to recover missing modalities from Gaussian noise, improving the distributional consistency of reconstructed representations. CIF-MMIN (Liu et al., 2024) leverages contrastive learning to extract modality-invariant semantics, reducing representation discrepancies induced by missing modalities. SDR-GNN (Fu et al., 2025) performs feature recovery and relational reasoning in the spectral domain, while HyperEF (Qiu et al., 2026) combines hypergraph-guided diffusion with evidential fusion to capture high-order semantic relations and predictive uncertainty. Despite these advances, jointly modeling *heavy-tailed recovery uncertainty*, *source-dependent evidence reliability*, and *local conflicts among multimodal predictions* remains underexplored.

To address these limitations, we propose **Lévy Flow and Uncertainty-Aware Fusion (LFUF)** for incomplete multimodal emotion recognition. LFUF first constructs context-aware priors for missing modalities and combines them with the available observations to guide conditional flow matching from an α-stable (i.e., Lévy-stable) source. We refer to this deterministic source-to-target transport as Lévy Flow for brevity. Unlike conventional Gaussian sources, the heavy-tailed Lévy stable source provides greater flexibility for modeling complex latent deviations caused by missing observations, while flow-based generation enables efficient recovery of semantically consistent representations. LFUF then treats the recovered modalities as complementary evidence sources and performs uncertainty-aware evidential fusion. Specifically, Jousselme consistency is used to estimate source reliability, and unreliable belief mass is reassigned to the uncertainty set to prevent noisy evidence from dominating the decision. Finally, PCR6 redistributes local conflicts among the involved hypotheses, yielding a robust emotion prediction under incomplete and potentially conflicting multimodal observations.

In summary, the main contributions are as follows:

- **Context-Guided Heavy-Tailed Flow Recovery.** We formulate missing-modality recovery as context-conditioned probabilistic transport, where speaker identities, conversational dependencies, and observed modalities jointly define the recovery condition. Unlike conventional Gaussian-source approaches, LFUF employs a symmetric α-stable source, i.e., Lévy stable source with a robust flow-matching objective, enabling more flexible modeling of irregular and heavy-tailed latent deviations caused by modality absence.
- **Finite-Step Shortcut Recovery.** We introduce a Shortcut-based recovery mechanism that learns finite-interval transport rather than relying solely on instantaneous vector fields. By enforcing consistency between one full update and two half updates through an EMA teacher, LFUF enables accurate recovery with only a few function evaluations while remaining compatible with heavy-tailed source distributions.
- **Reliability- and Conflict-Aware Evidential Fusion.** LFUF models complementary evidence at the unimodal, pairwise, and global levels instead of treating observed and recovered representations as equally reliable. Jousselme-consistency-based discounting transfers unreliable belief mass to ignorance, while hierarchical PCR6 redistributes local conflicts only among the responsible hypotheses, explicitly separating source unreliability from inter-source conflict.
- **Robust Incomplete Multimodal Emotion Recognition.** Experiments on IEMOCAP4, IEMOCAP6, and MELD under modality missing rates from 0.0 to 0.7 show consistent improvements over representative incomplete multimodal learning methods. Ablation and sensitivity studies further verify the effectiveness of heavy-tailed recovery, contextual conditioning, reliability discounting, and local conflict redistribution.

2 Related Work

Flow Matching: Diffusion models learn complex data distributions by progressively perturbing data with noise and approximating the reverse denoising process. Early work connected diffusion-based generation to nonequilibrium thermodynamics (Sohl-Dickstein et al., 2015), followed by de-

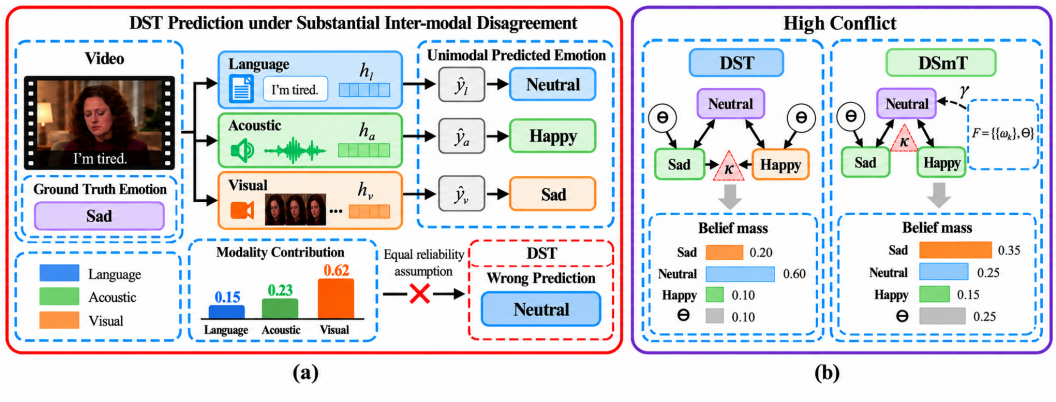


Figure 1: (a) DST prediction under substantial inter-modal disagreement, where equal modality reliability prevents effective modeling of differential contributions. (b) DST and DSmT predictions under severe evidence conflict. Unlike DST, DSmT uses a hyper-power set and local conflict redistribution to better accommodate conflicting evidence.

noising diffusion probabilistic models with Gaussian transitions (Ho et al., 2020). Score-based generative modeling further unified discrete diffusion and continuous-time stochastic differential equations (Song et al., 2021). Despite their strong performance in conditional generation and reconstruction, diffusion models typically require repeated network evaluations during sampling and rely on a prescribed stochastic process. Flow Matching instead learns the velocity field of a continuous normalizing flow by regressing the conditional velocity along a predefined probability path (Lipman et al., 2023; Tong et al., 2024). Its simulation-free training objective provides an efficient and flexible framework for transporting a tractable source distribution toward a conditional target distribution.

Lévy-Stable Source Distributions: Gaussian noise is widely used in generative modeling because of its analytical tractability and connection to Brownian motion. However, its exponentially decaying tails assign little probability mass to extreme deviations, limiting its ability to represent heterogeneous data with rare events, outliers, or imbalanced modes. Heavy-tailed distributions offer a more flexible alternative. In particular, the α-stable family generalizes the Gaussian distribution: $\alpha = 2$ recovers the Gaussian case, whereas $\alpha < 2$ yields polynomially decaying tails and, in general, infinite variance (Samorodnitsky & Taqqu, 1994; Nolan, 2020). As increments of a Lévy process, α-stable variables additionally induce discontinuous jumps and nonlocal transitions. Recent Lévy-based generative models suggest that heavy-tailed noise can improve tail coverage and better capture complex or imbalanced distributions (Yoon et al., 2023; Pandey et al., 2025; Shariatian et al., 2025).

Uncertainty Modeling: Uncertainty estimation assesses prediction reliability and is commonly decomposed into aleatoric uncertainty, arising from inherent data noise or ambiguity, and epistemic uncertainty, reflecting limited model or distributional knowledge (Kendall & Gal, 2017; Gawlikowski et al., 2023). Bayesian methods model distributions over parameters, whereas evidential deep learning places Dirichlet distributions over categorical probabilities and learns nonnegative evidence for competing hypotheses (Sensoy et al., 2018; Han et al., 2021). Recent studies extend evidential uncertainty modeling to incomplete multi-view and multimodal settings through divergence-based uncertainty quantification (Xue et al., 2025; Zhang et al., 2025) and uncertainty-aware evidence fusion coupled with missing-modality recovery (Qiu et al., 2026). Dempster–Shafer theory (DST) further represents incomplete knowledge by assigning belief masses to subsets of a frame of discernment (Dempster, 1967; Shafer, 1976), but its normalized combination rule can yield counterintuitive decisions under severe conflict (Zadeh, 1986; Yager, 1987). Evidence discounting and belief-assignment distances, such as the Jousselme distance, characterize source reliability and disagreement (Jousselme et al., 2001; Deng et al., 2004). Dezert–Smarandache theory (DSmT) generalizes DST to potentially nonexclusive hypotheses (Smarandache & Dezert, 2004), while proportional conflict redistribution rules, particularly PCR6, locally reallocate partial conflicts to the hypotheses that generate them rather than globally normalizing conflict (Smarandache & Dezert, 2006; Martin & Osswald, 2006), as illustrated in Fig. 1.

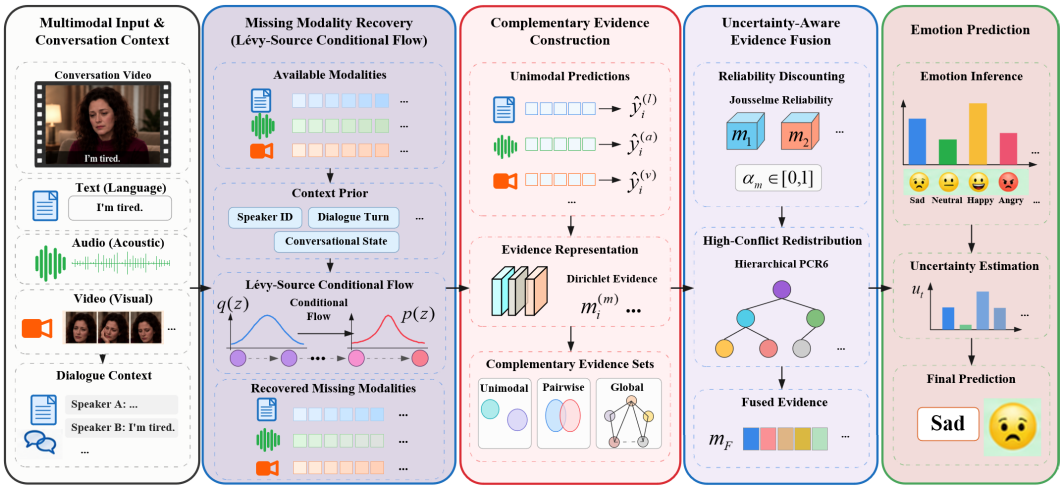


Figure 2: The framework integrates observed modalities with conversational priors to recover missing modalities via a Lévy-source Shortcut conditional flow, then constructs complementary evidence, applies Jousselme-based reliability discounting, and resolves local conflicts with hierarchical PCR6 for emotion prediction.

3 METHODOLOGY

3.1 PROBLEM FORMULATION AND LÉVY-SOURCE RECOVERY

In this section, we introduce LFUF, as illustrated in Fig. 2. Given a conversation $\mathcal{D} = \{u_i\}_{i=1}^{N}$ with modalities $\mathcal{M} = \{l, a, v\}$, let $\boldsymbol{x}_i^q$ denote the input of modality q and $r_i^q \in \{0, 1\}$ its availability. The observed and missing sets are $\mathcal{O}_i = \{q \in \mathcal{M} \mid r_i^q = 1\}, \overline{\mathcal{O}}_i = \mathcal{M} \setminus \mathcal{O}_i$, and modality encoders produce $\boldsymbol{h}_i^q = \mathcal{E}_q(\boldsymbol{x}_i^q)$. Our goal is to recover missing representations while accounting for their uncertainty before multimodal fusion. For each missing modality $q \in \overline{\mathcal{O}}_i$, we model $p\left(\boldsymbol{h}_i^q \mid \{\boldsymbol{h}_i^{q'} : q' \in \mathcal{O}_i\}, \boldsymbol{c}_i^q\right)$, where $\boldsymbol{c}_i^q = \Psi_q\left(\boldsymbol{q}_{\text{miss}}^q; \{\boldsymbol{h}_j^q, r_j^q, \pi_j, \tau_j\}_{j=1}^{N}\right)$, where $\boldsymbol{q}_{\text{miss}}^q$ is a learnable query, while π_j and τ_j denote speaker identity and turn index. Only available positions participate in contextual aggregation. The contextual encoder uses both preceding and subsequent available turns, placing LFUF in an offline bidirectional MERC setting without accessing the missing target representation. The resulting prior and observed modalities form the condition $\mathcal{C}_i^q$ without accessing the missing target (Ghosal et al., 2019; Hu et al., 2021; Lian et al., 2023).

Missing modalities can induce substantial non-Gaussian deviations in the latent space, which are difficult to capture using a light-tailed Gaussian source. To accommodate such irregular variations, we employ a heavy-tailed α-stable source, which provides broader support for atypical latent perturbations and improves the robustness of missing-modality recovery. Specifically, instead of a Gaussian source, we adopt a component-wise symmetric α-stable distribution:

$$Z_d^q \sim \mathcal{S}\alpha\mathcal{S}(\alpha_{\mathrm{L}}, \sigma_q), \qquad \mathbb{E}\left[e^{\mathrm{i}\omega Z_d^q}\right] = \exp\left(-\sigma_q^{\alpha_{\mathrm{L}}}|\omega|^{\alpha_{\mathrm{L}}}\right), \tag{1}$$

where $1 < \alpha_{\mathrm{L}} \leq 2$. A smaller α_{L} corresponds to heavier tails and therefore assigns greater probability mass to large latent deviations. The Gaussian distribution is recovered as the limiting boundary case: $\mathcal{S}\alpha\mathcal{S}(2, \sigma_q) = \mathcal{N}(0, 2\sigma_q^2)$. Importantly, the α-stable distribution is used only as the source measure of the deterministic conditional flow rather than as an explicit stochastic transition model. To improve numerical stability during training, extreme source samples are clipped using a sufficiently large threshold (Samorodnitsky & Taqqu, 1994; Nolan, 2020).

Given source sample $\widetilde{\boldsymbol{Z}}^q$ and target $\boldsymbol{H}^q$, we define

$$\boldsymbol{X}_t^q = (1-t)\widetilde{\boldsymbol{Z}}^q + t\boldsymbol{H}^q, \qquad \boldsymbol{U}^q = \boldsymbol{H}^q - \widetilde{\boldsymbol{Z}}^q. \tag{2}$$

The optimal conditional vector field satisfies

$$\boldsymbol{v}^*(\boldsymbol{x}, t, \mathcal{C}^q) = \mathbb{E}\left[\boldsymbol{U}^q \mid \boldsymbol{X}_t^q = \boldsymbol{x}, \mathcal{C}^q\right], \tag{3}$$

whose induced ODE transports the source toward the conditional target distribution (Lipman et al., 2023; Tong et al., 2024).

To reduce the influence of large velocity residuals, we adopt

$$\rho_{p,\epsilon}(e) = (e^2 + \epsilon^2)^{p/2} - \epsilon^p, \qquad 1 < p < \alpha_{\mathrm{L}}, \tag{4}$$

and optimize

$$\mathcal{L}_{\mathrm{LFM}} = \mathbb{E}\left[\rho_{p,\epsilon}\left(\boldsymbol{v}_\theta(\boldsymbol{X}_t^q, t, \mathcal{C}^q) - \boldsymbol{U}^q\right)\right]. \tag{5}$$

Source clipping is used solely for numerical stabilization and yields a bounded effective source. The condition $1 < p < \alpha_{\mathrm{L}}$ pertains to the underlying unclipped α-stable law and determines the exponent range of the robust penalty.

To reduce sampling cost, we further learn finite-step velocities using shortcut self-consistency (Frans et al., 2025). For flow map $\Phi_{t\to t+\delta}$,

$$\boldsymbol{s}^*(\boldsymbol{x}, t, \delta, \mathcal{C}) = \frac{\Phi_{t\to t+\delta}(\boldsymbol{x}) - \boldsymbol{x}}{\delta}. \tag{6}$$

An EMA teacher constructs the two-half-step target

$$\boldsymbol{v}_1 = \boldsymbol{s}_{\bar\theta}\left(\boldsymbol{x}_t, t, \frac{\delta}{2}, \mathcal{C}\right), \quad \boldsymbol{x}_{1/2} = \boldsymbol{x}_t + \frac{\delta}{2}\boldsymbol{v}_1, \quad \boldsymbol{v}_2 = \boldsymbol{s}_{\bar\theta}\left(\boldsymbol{x}_{1/2}, t + \frac{\delta}{2}, \frac{\delta}{2}, \mathcal{C}\right), \quad \boldsymbol{v}_{\mathrm{tar}} = \frac{\boldsymbol{v}_1 + \boldsymbol{v}_2}{2}. \tag{7}$$

with $\mathcal{L}_{\mathrm{SC}} = \mathbb{E}\left[\rho_{p,\epsilon}\left(\boldsymbol{s}_\theta(\boldsymbol{x}_t, t, \delta, \mathcal{C}) - \mathrm{sg}[\boldsymbol{v}_{\mathrm{tar}}]\right)\right]$. At inference, with $\Delta = 1/N_{\mathrm{FE}}$, $\widehat{\boldsymbol{h}}_{k+1}^q = \widehat{\boldsymbol{h}}_k^q + \Delta \boldsymbol{s}_{\bar\theta}\left(\widehat{\boldsymbol{h}}_k^q, k\Delta, \Delta, \mathcal{C}^q\right)$, followed by $\widehat{\boldsymbol{H}}^q = R_q(\widehat{\boldsymbol{h}}_{N_{\mathrm{FE}}}^q)$, $\mathcal{L}_{\mathrm{rec}} = \frac{1}{|\overline{\mathcal{O}}|}\sum_{q\in\overline{\mathcal{O}}} \mathrm{SmoothL1}(\widehat{\boldsymbol{H}}^q, \boldsymbol{H}^q)$.

3.2 Uncertainty-Aware Evidence Fusion

Missing-modality recovery inevitably introduces heterogeneous uncertainty across modalities, making direct feature fusion vulnerable to unreliable or conflicting recovered representations. Moreover, different evidence sources may exhibit substantially different levels of agreement. Treating all sources equally can therefore propagate, or even amplify, recovery errors during multimodal aggregation. To address this issue, we introduce an uncertainty-aware evidence fusion mechanism that explicitly estimates source reliability, discounts inconsistent evidence, and redistributes inter-source conflicts before producing the final prediction. We uniformly denote both observed and recovered features by $\overline{\boldsymbol{H}}^q$. Based on these features, we construct modality-specific, cross-modal, and globally fused representations as

$$\begin{cases} \boldsymbol{g}_q = \mathrm{Pool}\Big(\mathrm{SA}(\overline{\boldsymbol{H}}^q)\Big), \qquad \boldsymbol{g}_F = \mathrm{SA}([\boldsymbol{g}_l, \boldsymbol{g}_a, \boldsymbol{g}_v, \boldsymbol{t}_F])_{\boldsymbol{t}_F}, \\ \boldsymbol{g}_{qr} = \mathrm{Pool}\Big(\mathrm{SA}\Big[\mathrm{CA}(\overline{\boldsymbol{H}}^q, \overline{\boldsymbol{H}}^r) \| \, \mathrm{CA}(\overline{\boldsymbol{H}}^r, \overline{\boldsymbol{H}}^q)\Big]\Big), \qquad q < r, \end{cases} \tag{8}$$

where SA, CA, and Pool denote self-attention, cross-modal attention, and temporal pooling, respectively (Tsai et al., 2019; Hu et al., 2021). For each evidence source $s \in \mathcal{S}$, where $S = |\mathcal{S}|$, we first compute non-negative evidence and the corresponding Dirichlet parameters as

$$\boldsymbol{e}_s = \mathrm{Softplus}(W_s \boldsymbol{g}_s + \boldsymbol{b}_s), \quad \boldsymbol{\alpha}_s = \boldsymbol{e}_s + \boldsymbol{1}, \quad T_s = \sum_{k=1}^{K} e_{s,k} + K. \tag{9}$$

Following evidential deep learning (Sensoy et al., 2018; Han et al., 2021), the resulting evidence is mapped to a basic belief assignment (BBA) as

$$m_s(\{\omega_k\}) = \frac{e_{s,k}}{T_s}, \qquad m_s(\Theta) = \frac{K}{T_s}, \tag{10}$$

where $\Theta = \{\omega_1, \ldots, \omega_K\}$ denotes the frame of discernment, and $m_s(\Theta)$ represents the residual ignorance associated with source s.

To quantify inter-source consistency, we measure pairwise disagreement using the Jousselme distance (Jousselme et al., 2001):

$$d_J(m_s, m_r) = \sqrt{\frac{1}{2}(\boldsymbol{m}_s - \boldsymbol{m}_r)^\top D(\boldsymbol{m}_s - \boldsymbol{m}_r)}, \qquad D(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{11}$$

Based on the resulting pairwise similarities, we define the reliability of source s as

$$\gamma_s = \frac{\frac{1}{S-1}\sum_{r\neq s}[1 - d_J(m_s, m_r)]}{\max_j \frac{1}{S-1}\sum_{r\neq j}[1 - d_J(m_j, m_r)] + \varepsilon}, \tag{12}$$

and discount its BBA accordingly:

$$m_s^\gamma(\{\omega_k\}) = \gamma_s m_s(\{\omega_k\}), \qquad m_s^\gamma(\Theta) = 1 - \gamma_s + \gamma_s m_s(\Theta). \tag{13}$$

This reliability discounting transfers belief mass from potentially inconsistent singleton hypotheses to the ignorance set Θ, thereby reducing the influence of unreliable evidence sources during fusion.

The consensus-based reliability measure preserves broadly supported correct evidence and strengthens its relative contribution. For repeated or erroneous evidence that is inconsistent with the overall evidence, reliability discounting reduces its committed belief mass, thereby limiting its misleading influence and the conflict it induces before fusion.

The discounted evidence is subsequently fused using PCR6, which redistributes each elementary conflict only among the focal elements directly involved in that conflict (Smarandache & Dezert, 2006; Martin & Osswald, 2006). For $\boldsymbol{A} = (A_1, \ldots, A_S)$ satisfying $\bigcap_{s=1}^S A_s = \emptyset$, we define

$$\kappa(\boldsymbol{A}) = \prod_{s=1}^S m_s^\gamma(A_s), \qquad \Delta_s(\boldsymbol{A}) = \frac{[m_s^\gamma(A_s)]^2 \prod_{r\neq s} m_r^\gamma(A_r)}{\sum_{r=1}^S m_r^\gamma(A_r)}. \tag{14}$$

The resulting fused BBA is therefore

$$m_{\mathrm{PCR6}}(X) = m_\cap(X) + \sum_{\boldsymbol{A}:\, \cap_s A_s = \emptyset} \sum_{s:\, A_s = X} \Delta_s(\boldsymbol{A}), \qquad m_{\mathrm{PCR6}}(\emptyset) = 0. \tag{15}$$

Direct enumeration of all multi-source focal-element combinations can become computationally expensive as the number of evidence sources increases. Therefore, we recursively apply exact two-source PCR6 according to a fixed balanced fusion tree: $m_F = \mathrm{TreePCR6}\left(\{m_s^\gamma\}_{s=1}^S; \mathcal{T}\right)$. Because PCR6 is non-associative, the fusion tree $\mathcal{T}$ is fixed throughout both training and inference, thereby ensuring a deterministic and consistent fusion procedure.

Finally, the fused BBA is transformed into pignistic probabilities (Smets, 1990; Smets & Kennes, 1994):

$$p_k = m_F(\{\omega_k\}) + \frac{m_F(\Theta)}{K}. \tag{16}$$

The final emotion prediction is obtained as

$$\widehat{y} = \arg\max_{k\in\{1,\ldots,K\}} p_k. \tag{17}$$

The output is determined by the hypothesis with the largest pignistic probability. This formulation preserves uncertainty-aware evidence aggregation while enabling fine-grained continuous prediction.

3.3 Joint Learning Objective

Each evidence source is supervised by the expected Dirichlet negative log-likelihood

$$\mathcal{L}_{\mathrm{evi}} = \frac{1}{S}\sum_{s=1}^S \mathbb{E}\left[\sum_{k=1}^K y_k \left(\psi(T_s) - \psi(\alpha_{s,k})\right)\right], \tag{18}$$

where $\psi(\cdot)$ is the digamma function. Unsupported evidence is regularized toward the uniform Dirichlet prior:

$$\mathcal{L}_{\mathrm{KL}} = \frac{1}{S}\sum_{s=1}^S D_{\mathrm{KL}}\left[\mathrm{Dir}(\widetilde{\boldsymbol{\alpha}}_s) \| \,\mathrm{Dir}(\mathbf{1})\right], \qquad \widetilde{\boldsymbol{\alpha}}_s = \boldsymbol{\alpha}_s - \boldsymbol{y} \odot (\boldsymbol{\alpha}_s - \mathbf{1}). \tag{19}$$

The task loss is defined as

$$\mathcal{L}_{\text{task}} = -\frac{1}{N}\sum_{i=1}^{N}\log p_{i,y_i}, \tag{20}$$

where p_{i,y_i} denotes the pignistic probability assigned to the ground-truth emotion class of the i-th utterance.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_{\text{LFM}}\mathcal{L}_{\text{LFM}} + \lambda_{\text{SC}}\mathcal{L}_{\text{SC}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{evi}}\mathcal{L}_{\text{evi}} + \lambda_{\text{KL}}\mathcal{L}_{\text{KL}}. \tag{21}$$

Algorithm 1 summarizes LFUF. It recovers missing modalities from context-aware priors and Lévy sources, converts the completed representations into evidential opinions, discounts unreliable evidence via Jousselme distance, and fuses the resulting BBAs with Tree-PCR6 for final prediction. The theoretical analysis is provided in the Appendix A.

4 Experiments

4.1 Datasets and Experimental Settings

Algorithm 1: Optimization Procedure of LFUF

Input: $\{\boldsymbol{x}^q\}_{q\in\mathcal{M}}$, mask $\boldsymbol{r}$, label $\boldsymbol{y}$
Output: Prediction $\widehat{\boldsymbol{y}}$
$\boldsymbol{H}^q \leftarrow \mathcal{E}_q(\boldsymbol{x}^q)$, $\mathcal{O} \leftarrow \{q : r^q = 1\}$,
$\overline{\mathcal{O}} \leftarrow \mathcal{M} \setminus \mathcal{O}$;
for $q \in \overline{\mathcal{O}}$ **do**
  $\mathcal{C}^q \leftarrow$ ContextPrior$(\{\boldsymbol{H}^{q'}\}_{q'\in\mathcal{O}}, \boldsymbol{r}, \boldsymbol{\pi}, \boldsymbol{\tau})$;
  $\widetilde{\boldsymbol{Z}}^q \sim \mathcal{S}\alpha\mathcal{S}(\alpha_{\text{L}}, \sigma_q)$;
  $\widehat{\boldsymbol{H}}^q \leftarrow$ ShortcutRecover$(\widetilde{\boldsymbol{Z}}^q, \mathcal{C}^q; \boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$;
$\{\boldsymbol{g}_s\}_{s\in\mathcal{S}} \leftarrow$ EvidenceEncoder$(\{\overline{\boldsymbol{H}}^q\})$;
$\{m_s^\gamma\}_{s\in\mathcal{S}} \leftarrow$ JousselmeDiscount$\big($DirichletBBA$(\{\boldsymbol{g}_s\})\big)$;
$m_F \leftarrow$ TreePCR6$(\{m_s^\gamma\}_{s\in\mathcal{S}})$;
$\boldsymbol{p} \leftarrow$ Pignistic(m_F), $\widehat{\boldsymbol{y}} \leftarrow$ Decide$(\boldsymbol{p})$;
$\boldsymbol{\vartheta} \leftarrow \boldsymbol{\vartheta} - \eta\nabla_{\boldsymbol{\vartheta}}\mathcal{L}$;
$\bar{\boldsymbol{\theta}} \leftarrow \mu\bar{\boldsymbol{\theta}} + (1-\mu)\boldsymbol{\theta}$;

Experiments are conducted on IEMOCAP (Busso et al., 2008) and MELD (Poria et al., 2019). IEMOCAP contains 151 dyadic dialogues and 7,433 multimodal utterances from ten actors, and we evaluate both its four-class (IEMOCAP4) and six-class (IEMOCAP6) settings. MELD contains 1,432 multiparty dialogues and 13,708 utterances from 304 speakers across seven emotion categories. We adopt utterance-level random modality missingness, defined as $\kappa = 1 - \frac{\sum_{i=1}^{N} m_i}{NM}$ with $\kappa \in 0.0, 0.1, \ldots, 0.7$, while retaining at least one observed modality per utterance. For IEMOCAP, the first four sessions are used for training and the fifth for evaluation; MELD follows its official split. The model is jointly trained in a single stage using AdamW with a learning rate of 5×10^{-5} and batch sizes of 8 and 64 for IEMOCAP and MELD, respectively. We set the symmetric α-stable source to $\alpha = 1.8$ and $\sigma = 1.0$, with a component-wise clipping threshold of 10σ. Under the default setting, approximately 0.32% of scalar source samples exceed this threshold. The robust penalty $\rho_{p,\epsilon}$ to $p = 1.5$ and $\epsilon = 10^{-3}$. The loss weights are fixed as $(\lambda_{\text{LFM}}, \lambda_{\text{SC}}, \lambda_{\text{rec}}, \lambda_{\text{evi}}, \lambda_{\text{KL}}) = (0.5, 0.5, 0.1, 0.1, 0.1)$ for all experiments. Flow recovery uses four function evaluations. All experiments are conducted on a single NVIDIA L40S GPU with 90 GB of memory.

4.2 Comparison with State-of-the-Art Methods

We compare LFUF with six representative methods for incomplete multimodal emotion recognition, spanning modality reconstruction, graph-based modeling, contrastive learning, generative recovery, and uncertainty-aware fusion. MMIN (Zhao et al., 2021) learns robust joint representations and reconstructs missing modalities from available observations. GCNet (Lian et al., 2023) models speaker and temporal dependencies with conversational graphs while jointly optimizing modality reconstruction and emotion recognition. IMDer (Wang et al., 2023) employs conditional score-based diffusion to recover missing representations under the guidance of observed modalities. CIF-MMIN (Liu et al., 2024) uses contrastive learning to reduce cross-modal discrepancies and improve missing-modality reconstruction. SDR-GNN (Fu et al., 2025) captures speaker and contextual relations through interaction graphs and spectral multi-frequency aggregation. HyperEF (Qiu et al., 2026) combines hypergraph-conditioned diffusion with evidential fusion to capture high-order semantic dependencies and quantify uncertainty under incomplete observations. We followed the experimental setup of HyperEF (Qiu et al., 2026).

Table 1: Accuracy (%) across different modality missing rates. For LFUF, the stability parameter is set to $\alpha = 1.8$, and the robust penalty exponent is set to $p = 1.5$. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Method	Venue	Missing Rate							
			0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
IEMOCAP4	MMIN (Zhao et al., 2021)	ACL'21	74.9	71.8	69.4	66.3	63.3	60.5	57.5	55.4
	GCNet (Lian et al., 2023)	TPAMI'23	78.4	77.5	77.3	76.2	75.1	73.8	71.9	71.4
	IMDer (Wang et al., 2023)	NeurIPS'23	74.3	73.5	69.5	64.1	59.2	64.5	46.9	44.2
	CIF-MMIN (Liu et al., 2024)	TAFFC'24	79.3	77.9	77.4	77.3	77.1	76.0	74.7	73.6
	SDR-GNN (Fu et al., 2025)	KBS'25	79.6	78.6	78.1	77.5	77.1	75.8	75.0	74.4
	HyperEF (Qiu et al., 2026)	CVPR'26	<u>82.9</u>	<u>82.1</u>	<u>79.1</u>	<u>80.3</u>	<u>77.6</u>	<u>78.2</u>	<u>77.7</u>	<u>77.0</u>
	LFUF (Ours)	—	**85.9**	**85.4**	**85.2**	**84.3**	**83.6**	**82.4**	**80.3**	**79.7**
IEMOCAP6	MMIN (Zhao et al., 2021)	ACL'21	57.0	53.9	51.5	48.4	45.6	42.8	40.2	37.8
	GCNet (Lian et al., 2023)	TPAMI'23	58.6	58.5	57.6	57.1	56.1	54.4	53.6	53.5
	IMDer (Wang et al., 2023)	NeurIPS'23	57.9	52.9	53.5	46.0	49.9	39.6	31.1	28.2
	CIF-MMIN (Liu et al., 2024)	TAFFC'24	60.5	58.8	57.3	56.8	56.2	55.7	55.1	54.3
	SDR-GNN (Fu et al., 2025)	KBS'25	61.2	60.9	59.8	59.5	59.2	57.4	55.5	55.3
	HyperEF (Qiu et al., 2026)	CVPR'26	<u>64.1</u>	<u>63.2</u>	<u>61.8</u>	<u>61.1</u>	<u>60.1</u>	<u>59.8</u>	<u>59.5</u>	<u>58.5</u>
	LFUF (Ours)	—	**68.2**	**66.4**	**65.1**	**64.5**	**63.4**	**62.2**	**61.1**	**58.7**
MELD	MMIN (Zhao et al., 2021)	ACL'21	64.9	59.8	58.8	55.8	54.8	53.7	51.3	49.3
	GCNet (Lian et al., 2023)	TPAMI'23	65.5	63.3	62.2	59.7	58.3	57.2	55.2	53.6
	IMDer (Wang et al., 2023)	NeurIPS'23	67.1	66.1	62.6	61.4	60.5	58.4	57.0	55.3
	CIF-MMIN (Liu et al., 2024)	TAFFC'24	65.8	65.3	62.7	61.2	59.6	58.0	56.4	55.3
	SDR-GNN (Fu et al., 2025)	KBS'25	63.8	62.4	60.9	57.8	57.4	56.6	54.4	52.9
	HyperEF (Qiu et al., 2026)	CVPR'26	<u>69.3</u>	<u>67.2</u>	<u>64.8</u>	<u>63.8</u>	<u>61.8</u>	<u>60.4</u>	<u>58.6</u>	<u>57.1</u>
	LFUF (Ours)	—	**69.6**	**68.3**	**66.6**	**64.8**	**62.3**	**61.2**	**58.8**	**57.2**

We evaluate all methods on IEMOCAP4, IEMOCAP6, and MELD with random missing rates from 0.0 to 0.7. For each utterance, modalities are randomly dropped while retaining at least one observed modality. The accuracy results are reported in Table 1. The sensitivity of LFUF to the stability parameter α is further examined in Appendix B. Additional controlled comparisons that separately examine reliability discounting and evidence combination rules are provided in Appendix C. The effect of Shortcut self-consistency and its stability across different inference budgets are further analyzed in Appendix D.

4.3 Results and Analyses

4.3.1 Main Results

As shown in Table 1, LFUF consistently outperforms the competing methods across all datasets and missing-rate settings. These results demonstrate that LFUF generalizes effectively across different emotion categories and conversational structures. Although performance gradually degrades as the missing rate increases, LFUF retains a consistent advantage under incomplete-modality conditions, highlighting the effectiveness of integrating Lévy flow recovery with uncertainty-aware evidence fusion. Under severe modality absence, the narrowing performance margin further suggests that recovery quality is fundamentally limited by the amount of available observed information.

4.3.2 Visualization of Recovered Representations

As shown in Fig. 3, the dark and light points denote the original and recovered features, respectively. IMDer initializes recovery from Gaussian noise and progressively reconstructs the missing modalities through a reverse SDE, whereas LFUF adopts a heavy-tailed Lévy source and conditional flow matching. The heavy-tailed Lévy distribution provides broader coverage of irregular latent deviations, while the learned conditional velocity field transports the sampled representations toward the target modality distribution. On both IEMOCAP and MELD, the representations recovered by LFUF show greater overlap with the original features and exhibit more consistent distributional structures. These visualizations provide qualitative evidence that Lévy flow recovery more effectively preserves modality-specific semantic structure during missing-modality reconstruction.

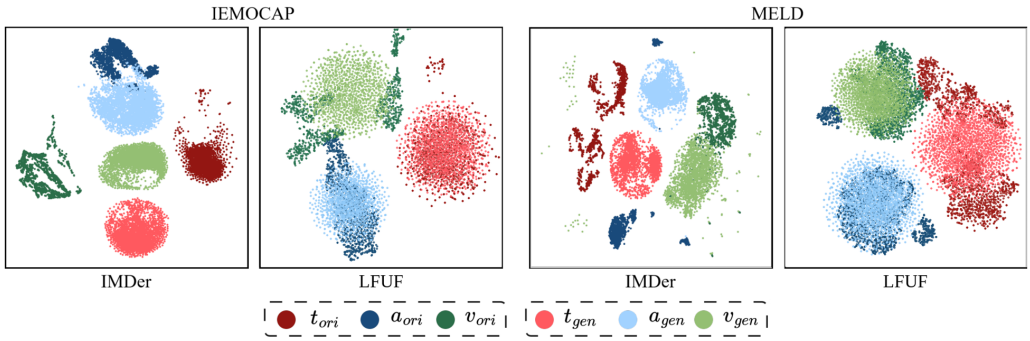


Figure 3: t-SNE visualization comparing the modality recovery performance of IMDer and LFUF on IEMOCAP and MELD.

Table 2: Ablation results on IEMOCAP and MELD under varying modality missing rates. Accuracy and F1 scores are reported as percentages.

Dataset	Variant	0.3		0.5		0.7	
		Acc	F1	Acc	F1	Acc	F1
IEMOCAP4	w/o Context Prior	83.3	83.0	81.4	81.0	78.2	78.0
	w/o Recovery	83.4	83.1	80.9	80.5	75.8	75.6
	w/o Modality Interaction	84.1	84.0	80.3	80.1	77.8	77.5
	LFUF (Full)	**84.3**	**84.1**	**82.4**	**82.2**	**79.7**	**79.4**
MELD	w/o Context Prior	62.4	58.4	59.1	54.3	56.5	49.8
	w/o Recovery	62.4	56.9	59.3	54.0	56.9	48.8
	w/o Modality Interaction	53.5	39.8	51.7	36.6	51.0	35.1
	LFUF (Full)	**64.8**	**62.8**	**61.2**	**58.3**	**57.2**	**50.4**

4.3.3 Ablation Study

To assess the contributions of the key components, we conduct ablation experiments under the same training protocol across different datasets and modality missing rates. As shown in Table 2, removing the contextual prior consistently degrades performance, highlighting its importance in providing conversational constraints for missing-modality recovery. Disabling modality recovery likewise reduces recognition performance, confirming the benefit of recovering missing semantic information. The largest performance degradation occurs when modality interaction is removed and only unimodal evidence sources are retained, indicating that pairwise cross-modal and global representations provide complementary information beyond that available from individual modalities. Overall, these results demonstrate that contextual priors, modality recovery, and modality interaction each contribute to the robustness and recognition performance of LFUF.

5 Conclusion

This work presents LFUF, a unified framework for incomplete multimodal emotion recognition in conversations. LFUF leverages contextual priors and available observations to guide Lévy-source conditional flow matching, enabling the recovery of missing modalities under complex latent deviations. The recovered modalities are subsequently organized into complementary evidence sources, with source reliability and inter-source conflicts handled through Jousselme-distance-based discounting and PCR6 fusion. Experiments on IEMOCAP and MELD demonstrate consistent performance improvements across a range of modality missing rates. Further analyses of the stability parameter, evidence combination rules, and model components support the effectiveness of the heavy-tailed source distribution, context-guided recovery, and multi-level evidence fusion. Future work will explore adaptive source distributions and more diverse, realistic patterns of modality missingness.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript organization, translation, language refinement, reference formatting, and code debugging, as well as to provide feedback on theoretical presentation and experimental design. All references, AI-assisted code, equations, experimental results, and scientific claims were independently reviewed and validated by the authors. Generative AI was not used to create or modify experimental datasets or reported measurements. The authors take full responsibility for the final content of this work.

REFERENCES

Jonathan T. Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4331–4339, 2019. doi: 10.1109/CVPR.2019.00446.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335–359, 2008. doi: 10.1007/s10579-008-9076-6.

Arthur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967. doi: 10.1214/aoms/1177698950.

Yong Deng, Wenkang Shi, Zhenfu Zhu, and Qi Liu. Combining belief functions based on distance of evidence. *Decision Support Systems*, 38(3):489–493, 2004. doi: 10.1016/j.dss.2004.04.015.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations*, 2025.

Fangze Fu, Wei Ai, Fan Yang, Yuntao Shou, Tao Meng, and Keqin Li. SDR-GNN: Spectral domain reconstruction graph neural network for incomplete multimodal learning in conversational emotion recognition. *Knowledge-Based Systems*, 309:112825, 2025. doi: 10.1016/j.knosys.2024.112825.

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56:1513–1589, 2023. doi: 10.1007/s10462-023-10562-9.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 154–164, 2019. doi: 10.18653/v1/D19-1015.

Zirun Guo, Tao Jin, and Zhou Zhao. Multimodal prompt learning with missing modalities for sentiment analysis and emotion recognition. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 1726–1736, 2024. doi: 10.18653/v1/2024.acl-long.94.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *International Conference on Learning Representations*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Jingwen Hu, Yuchen Liu, Jinming Zhao, and Qin Jin. MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 5666–5675, 2021. doi: 10.18653/v1/2021.acl-long.440.

Anne-Laure Jousselme, Dominic Grenier, and Éloi Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2(2):91–101, 2001. doi: 10.1016/S1566-2535(01)00026-4.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Zheng Lian, Lan Chen, Licai Sun, Bin Liu, and Jianhua Tao. GCNet: Graph completion network for incomplete multimodal learning in conversation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8419–8432, 2023. doi: 10.1109/TPAMI.2023.3234553.

Ronghao Lin and Haifeng Hu. MissModal: Increasing robustness to missing modality in multimodal sentiment analysis. *Transactions of the Association for Computational Linguistics*, 11:1686–1702, 2023. doi: 10.1162/tacl_a_00628.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Rui Liu, Haolin Zuo, Zheng Lian, Björn W. Schuller, and Haizhou Li. Contrastive learning based modality-invariant feature acquisition for robust multimodal emotion recognition with missing modalities. *IEEE Transactions on Affective Computing*, 15(4):1856–1873, 2024. doi: 10.1109/TAFFC.2024.3378570.

Mengmeng Ma, Jian Ren, Long Zhao, Davide Testuggine, and Xi Peng. Are multimodal transformers robust to missing modality? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18177–18186, 2022.

Arnaud Martin and Christophe Osswald. A new generalization of the proportional conflict redistribution rule stable in terms of decision. In Florentin Smarandache and Jean Dezert (eds.), *Advances and Applications of DSmT for Information Fusion: Collected Works, Volume 2*, pp. 69–88. American Research Press, 2006. URL https://fs.unm.edu/ANewGeneralizationOfTheProportionalConflict.pdf.

David Mercier, Benjamin Quost, and Thierry Denœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258, 2008. doi: 10.1016/j.inffus.2006.08.001.

John P. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer, 2020. doi: 10.1007/978-3-030-52915-4.

Kushagra Pandey, Jaideep Pathak, Yilun Xu, Stephan Mandt, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Heavy-tailed diffusion models. In *International Conference on Learning Representations*, 2025.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 527–536, 2019. doi: 10.18653/v1/P19-1050.

Xihang Qiu, Yuhao Fang, Qing Zhou, Bin Zhai, Jialong Hong, Wanpeng Zhang, Yao Lu, Ye Zhang, and Chun Li. Beyond missing modalities: Hypergraph conditioned diffusion for uncertainty-aware multimodal emotion recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22953–22963, 2026.

Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall/CRC, 1994. doi: 10.1201/9780203738818.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

Dario Shariatian, Umut Simsekli, and Alain Oliviero Durmus. Denoising lévy probabilistic models. In *International Conference on Learning Representations*, 2025.

Umut Simsekli. Fractional langevin monte carlo: Exploring lévy driven stochastic differential equations for markov chain monte carlo. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3200–3209, 2017.

Florentin Smarandache and Jean Dezert (eds.). *Advances and Applications of DSmT for Information Fusion: Collected Works, Volume 1*. American Research Press, 2004. URL https://fs.unm.edu/DSmT-book1.pdf.

Florentin Smarandache and Jean Dezert. Proportional conflict redistribution rules for information fusion. In Florentin Smarandache and Jean Dezert (eds.), *Advances and Applications of DSmT for Information Fusion: Collected Works, Volume 2*, pp. 3–68. American Research Press, 2006. URL https://fs.unm.edu/DSmT-book2.pdf.

Philippe Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990. doi: 10.1109/34.55104.

Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66(2): 191–234, 1994. doi: 10.1016/0004-3702(94)90026-4.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J. Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6558–6569, 2019. doi: 10.18653/v1/P19-1656.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009. doi: 10.1007/978-3-540-71050-9.

Yuanzhi Wang, Yong Li, and Zhen Cui. Incomplete multimodality-diffused emotion recognition. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0748.

ChengYan Wu, Yiqiang Cai, Yang Liu, Pengxu Zhu, Yun Xue, Ziwei Gong, Julia Hirschberg, and Bolei Ma. Multimodal emotion recognition in conversations: A survey of methods, trends, challenges and prospects. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 6257–6274, 2025. doi: 10.18653/v1/2025.findings-emnlp.332.

Zhipeng Xue, Yan Zhang, Ming Li, Chun Li, Yue Liu, and Fei Richard Yu. Uncertainty quantification for incomplete multi-view data using divergence measures. *IEEE Transactions on Image Processing*, 34:4328–4342, 2025. doi: 10.1109/TIP.2025.3579987.

Ronald R. Yager. On the dempster–shafer framework and new combination rules. *Information Sciences*, 41(2):93–137, 1987. doi: 10.1016/0020-0255(87)90007-7.

Eun Bi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim. Score-based generative models with lévy processes. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-1772.

Lotfi A. Zadeh. A simple view of the dempster–shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, 7(2):85–90, 1986. doi: 10.1609/aimag.v7i2.542.

Yan Zhang, Ming Li, Chun Li, Zhaoxia Liu, Ye Zhang, and Fei Richard Yu. Uncertainty quantification via hölder divergence for multi-view representation learning. *IEEE Transactions on Multimedia*, 27:8263–8275, 2025. doi: 10.1109/TMM.2025.3604966.

Jinming Zhao, Ruichen Li, and Qin Jin. Missing modality imagination network for emotion recognition with uncertain missing modalities. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 2608–2618, 2021. doi: 10.18653/v1/2021.acl-long.203.

A Theoretical Analysis

We provide theoretical support for the two key components of LFUF: heavy-tailed finite-step recovery and reliability-aware evidential fusion. The analysis is conditioned on a fixed missing modality q and its associated context $\mathcal{C}^q$; the superscript q is omitted whenever no ambiguity arises. For a random vector $\boldsymbol{X}$, we define $\|\boldsymbol{X}\|_{L^p} = (\mathbb{E}\|\boldsymbol{X}\|_2^p)^{1/p}$.

For clarity, the robust penalty applied to a vector residual $\boldsymbol{e}$ is defined as

$$\rho_{p,\epsilon}(\boldsymbol{e}) = \left(\|\boldsymbol{e}\|_2^2 + \epsilon^2\right)^{p/2} - \epsilon^p, \qquad 1 < p < \alpha_{\mathrm{L}}. \tag{22}$$

This penalty belongs to the family of generalized robust losses designed to reduce the influence of large residuals (Barron, 2019).

A.1 Assumptions

Assumption 1 (Heavy-tail-compatible source and integrability). Let the raw source vector be $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top$, with each component following a symmetric stable distribution,

$$Z_j \sim \mathcal{S}\alpha\mathcal{S}(\alpha_{\mathrm{L}}, \sigma), \qquad 1 < p < \alpha_{\mathrm{L}} \leq 2. \tag{23}$$

When source clipping is employed, the effective source entering the model is

$$\widetilde{\boldsymbol{Z}} = T_B(\boldsymbol{Z}), \qquad T_B(z) = \mathrm{sign}(z)\min\{|z|, B\}, \tag{24}$$

where T_B is applied component-wise. Without clipping, T_B is understood as the identity map. The conditional target representation satisfies

$$\mathbb{E}\left[\|\boldsymbol{H}\|_2^p \mid \mathcal{C}\right] < \infty. \tag{25}$$

For $0 < \alpha_{\mathrm{L}} < 2$, an α_{L}-stable random variable has finite absolute moments of order $p < \alpha_{\mathrm{L}}$ (Samorodnitsky & Taqqu, 1994; Nolan, 2020). For $\alpha_{\mathrm{L}} = 2$, the distribution is Gaussian and therefore has finite moments of all positive orders. Since $d < \infty$, finite component-wise p-moments imply $\mathbb{E}\|\boldsymbol{Z}\|_2^p < \infty$. Clipping, when used, additionally makes $\widetilde{\boldsymbol{Z}}$ bounded.

Assumption 2 (Regular conditional transport). For every admissible condition $\mathcal{C}$, there exists a conditional flow map $\Phi_{t\to t+\delta}^{\mathcal{C}}$ whose terminal push-forward matches the conditional target distribution. Define the corresponding exact finite-step displacement field as

$$\boldsymbol{s}^\dagger(\boldsymbol{x}, t, \delta, \mathcal{C}) = \frac{\Phi_{t\to t+\delta}^{\mathcal{C}}(\boldsymbol{x}) - \boldsymbol{x}}{\delta}. \tag{26}$$

For every inference step size considered, $\boldsymbol{s}^\dagger$ is uniformly Lipschitz in its state argument:

$$\left\|\boldsymbol{s}^\dagger(\boldsymbol{x}, t, \delta, \mathcal{C}) - \boldsymbol{s}^\dagger(\boldsymbol{y}, t, \delta, \mathcal{C})\right\|_2 \leq L_s\|\boldsymbol{x} - \boldsymbol{y}\|_2. \tag{27}$$

Flow matching provides a simulation-free construction of vector fields associated with prescribed probability paths (Lipman et al., 2023). Generalized conditional flow matching further accommodates non-Gaussian source distributions without requiring explicit evaluation of the source density (Tong et al., 2024). Equation equation 27 is imposed as an explicit regularity condition rather than derived from the training objective.

Assumption 3 (Rollout shortcut control and transport anchoring). Let $\widehat{\boldsymbol{X}}_k$ denote the states visited by the learned shortcut model during inference, with $t_k = k\Delta$ and $\Delta = 1/N_{\mathrm{FE}}$. For every $k = 0, \ldots, N_{\mathrm{FE}} - 1$, assume

$$\mathbb{E}\left[\rho_{p,\epsilon}\left(\boldsymbol{s}_{\bar{\theta}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) - \boldsymbol{v}_{\mathrm{tar}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C})\right)\right] \le \ell_{\mathrm{SC}}, \tag{28}$$

and

$$\left\|\boldsymbol{v}_{\mathrm{tar}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) - \boldsymbol{s}^{\dagger}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C})\right\|_{L^p} \le \eta_{\mathrm{T}}. \tag{29}$$

The first condition requires shortcut consistency to remain controlled along inference trajectories rather than only over the training distribution. The second imposes an explicit transport-anchoring condition. This distinction is important because standard flow-matching equivalence is established through quadratic vector-field regression (Lipman et al., 2023; Tong et al., 2024), whereas LFUF intentionally uses the robust penalty in Eq. equation 22. Accordingly, η_{T} captures both residual teacher approximation error and potential anchoring bias introduced by robust regression. The finite-step self-consistency construction is motivated by Shortcut Models (Frans et al., 2025).

Assumption 4 (Restricted focal structure and fixed fusion tree). Let $S \ge 2$. Each evidence source has focal elements restricted to

$$\mathcal{F} = \{\{\omega_1\}, \ldots, \{\omega_K\}, \Theta\}, \tag{30}$$

and binary PCR6 is recursively applied using the same fixed tree $\mathcal{T}$ during both training and inference.

The Dirichlet-to-BBA mapping employed by LFUF automatically satisfies

$$\sum_{k=1}^{K} m_s(\{\omega_k\}) + m_s(\Theta) = \frac{\sum_k e_{s,k} + K}{T_s} = 1, \tag{31}$$

so BBA normalization does not require an additional assumption. Fixing $\mathcal{T}$ is necessary because PCR6 is generally non-associative (Martin & Osswald, 2006).

A.2 Finite-Step Stability of Heavy-Tailed Recovery

Theorem 1 (Heavy-tail-compatible finite-step recovery bound). Under Assumptions 1–3, let $\mu_0^{\mathcal{C}}$ denote the distribution of the effective source $\widetilde{\boldsymbol{Z}}$, and let $\nu_{\mathcal{C}}$ denote the terminal distribution of the exact conditional flow. Let $\widehat{\mu}_1^{\mathcal{C}}$ denote the endpoint distribution produced using N_{FE} shortcut evaluations.

Then the robust recovery objective is finite under the stated moment conditions. Moreover, defining

$$\varepsilon_s = (\ell_{\mathrm{SC}} + \epsilon^p)^{1/p} + \eta_{\mathrm{T}}, \tag{32}$$

we obtain

$$W_p\left(\widehat{\mu}_1^{\mathcal{C}}, \nu_{\mathcal{C}}\right) \le \frac{(1 + L_s\Delta)^{N_{\mathrm{FE}}} - 1}{L_s}\varepsilon_s \le \frac{e^{L_s} - 1}{L_s}\varepsilon_s, \tag{33}$$

with the continuous interpretation $W_p \le \varepsilon_s$ when $L_s = 0$.

Proof. For the unclipped source, $p < \alpha_{\mathrm{L}}$ implies

$$\mathbb{E}|Z_j|^p < \infty \tag{34}$$

for every component (Samorodnitsky & Taqqu, 1994; Nolan, 2020). Since $1 < p \le 2$, $\|\boldsymbol{Z}\|_2 \le \|\boldsymbol{Z}\|_p$, and therefore

$$\mathbb{E}\|\boldsymbol{Z}\|_2^p \le \sum_{j=1}^{d} \mathbb{E}|Z_j|^p < \infty. \tag{35}$$

If clipping is applied, $\|\widetilde{\boldsymbol{Z}}\|_2 \le \sqrt{d}\, B$, so the same conclusion follows immediately.

By Assumption 1, $\|\boldsymbol{H}\|_{L^p} < \infty$. Minkowski's inequality therefore gives

$$\|\boldsymbol{H} - \widetilde{\boldsymbol{Z}}\|_{L^p} \le \|\boldsymbol{H}\|_{L^p} + \|\widetilde{\boldsymbol{Z}}\|_{L^p} < \infty. \tag{36}$$

Because $1 < p < \alpha_{\mathrm{L}} \le 2$, we have $0 < p/2 \le 1$. For $a, b \ge 0$ and $0 < r \le 1$, $(a+b)^r \le a^r + b^r$. Consequently,

$$0 \le \rho_{p,\epsilon}(\boldsymbol{e}) \le \|\boldsymbol{e}\|_2^p. \tag{37}$$

Hence, under the finite p-moment conditions above, the robust flow objective is finite.

Conversely,

$$\|\boldsymbol{e}\|_2^p \le \rho_{p,\epsilon}(\boldsymbol{e}) + \epsilon^p. \tag{38}$$

Applying Eq. equation 38 to Assumption 3 yields

$$\left\| \boldsymbol{s}_{\bar{\theta}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) - \boldsymbol{v}_{\mathrm{tar}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) \right\|_{L^p} \le \left(\ell_{\mathrm{SC}} + \epsilon^p\right)^{1/p}. \tag{39}$$

Combining this inequality with Eq. equation 29 and applying Minkowski's inequality gives

$$\begin{aligned} &\left\| \boldsymbol{s}_{\bar{\theta}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) - \boldsymbol{s}^\dagger(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}) \right\|_{L^p} \\ &\le \left(\ell_{\mathrm{SC}} + \epsilon^p\right)^{1/p} + \eta_{\mathrm{T}} = \varepsilon_s. \end{aligned} \tag{40}$$

Couple the learned and exact trajectories using the same effective source:

$$\widehat{\boldsymbol{X}}_{k+1} = \widehat{\boldsymbol{X}}_k + \Delta \boldsymbol{s}_{\bar{\theta}}(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C}), \tag{41}$$

$$\boldsymbol{X}^\dagger_{k+1} = \boldsymbol{X}^\dagger_k + \Delta \boldsymbol{s}^\dagger(\boldsymbol{X}^\dagger_k, t_k, \Delta, \mathcal{C}), \tag{42}$$

where $\widehat{\boldsymbol{X}}_0 = \boldsymbol{X}^\dagger_0 = \widetilde{\boldsymbol{Z}}$. By Eq. equation 26, the second recursion follows the exact flow map at the selected grid points.

Define

$$E_k = \left\| \widehat{\boldsymbol{X}}_k - \boldsymbol{X}^\dagger_k \right\|_{L^p}. \tag{43}$$

Adding and subtracting $\boldsymbol{s}^\dagger(\widehat{\boldsymbol{X}}_k, t_k, \Delta, \mathcal{C})$ and applying Minkowski's inequality, Eq. equation 40, and Assumption 2, we obtain

$$E_{k+1} \le (1 + L_s \Delta) E_k + \Delta \varepsilon_s. \tag{44}$$

Since $E_0 = 0$,

$$\begin{aligned} E_{N_{\mathrm{FE}}} &\le \Delta \varepsilon_s \sum_{j=0}^{N_{\mathrm{FE}}-1} (1 + L_s \Delta)^j \\ &= \frac{(1 + L_s \Delta)^{N_{\mathrm{FE}}} - 1}{L_s} \varepsilon_s. \end{aligned} \tag{45}$$

Using $\Delta = 1/N_{\mathrm{FE}}$ and $(1 + L_s/N_{\mathrm{FE}})^{N_{\mathrm{FE}}} \le e^{L_s}$, we further obtain

$$E_{N_{\mathrm{FE}}} \le \frac{e^{L_s} - 1}{L_s} \varepsilon_s. \tag{46}$$

The constructed endpoint pair defines a valid coupling between $\widehat{\mu}_1^{\mathcal{C}}$ and $\nu_{\mathcal{C}}$. Therefore, by the definition of W_p,

$$W_p\left(\widehat{\mu}_1^{\mathcal{C}}, \nu_{\mathcal{C}}\right) \le E_{N_{\mathrm{FE}}}, \tag{47}$$

which establishes Eq. equation 33. □

Remark 1 (Output readout). If a deterministic L_R-Lipschitz readout R_q is applied to the terminal flow state, then

$$W_p\left((R_q)_{\#}\widehat{\mu}_1^{\mathcal{C}}, (R_q)_{\#}\nu_{\mathcal{C}}\right) \le L_R W_p\left(\widehat{\mu}_1^{\mathcal{C}}, \nu_{\mathcal{C}}\right) \tag{48}$$

(Villani, 2009). This result concerns the push-forward target $(R_q)_{\#}\nu_{\mathcal{C}}$. If the reconstruction loss is instead defined directly against the pre-readout target $\boldsymbol{H}$, then R_q should either be the identity in this theoretical abstraction or the target space should be defined consistently before applying Eq. equation 48.

Remark 2 (Role of robust regression). Theorem 1 does not assert that minimizing the robust objective alone identifies the exact flow-matching vector field. The term η_{T} explicitly accounts for the discrepancy between the robustly learned finite-step target and the exact transport field. This distinction is necessary because the conditional-mean characterization in standard flow matching follows from quadratic regression (Lipman et al., 2023; Tong et al., 2024), whereas LFUF deliberately adopts a p-order robust surrogate.

A.3 Reliability Contraction and Local Conflict Control

Let m_{V} denote the vacuous BBA,

$$m_{\mathrm{V}}(\Theta) = 1, \qquad m_{\mathrm{V}}(A) = 0 \quad \text{for } A \neq \Theta. \tag{49}$$

Reliability discounting weakens a belief source by transferring part of its committed mass to ignorance (Mercier et al., 2008).

Theorem 2 (Reliability contraction and conflict conservation under fixed Tree-PCR6). Under Assumption 4, embed each BBA in the common vector space indexed by 2^{Θ}, assigning zero mass to focal elements outside $\mathcal{F}$. Let $m_s^{\gamma_s}$ denote the discounted BBA of source s. Then

$$d_J\left(m_s^{\gamma_s}, m_{\mathrm{V}}\right) = \gamma_s d_J\left(m_s, m_{\mathrm{V}}\right), \tag{50}$$

and

$$\mathrm{BetP}\left(m_s^{\gamma_s}\right) = \gamma_s \,\mathrm{BetP}(m_s) + (1-\gamma_s)\boldsymbol{u}, \qquad \boldsymbol{u} = \frac{1}{K}\mathbf{1}. \tag{51}$$

Consequently, for any norm,

$$\left\|\mathrm{BetP}(m_s^{\gamma_s}) - \boldsymbol{u}\right\| = \gamma_s \left\|\mathrm{BetP}(m_s) - \boldsymbol{u}\right\|. \tag{52}$$

Moreover, for two sources s and r, their conjunctive singleton conflict,

$$\kappa_{sr} = \sum_{k \neq j} m_s(\{\omega_k\}) m_r(\{\omega_j\}), \tag{53}$$

satisfies, after discounting,

$$\kappa_{sr}^{\gamma} = \gamma_s \gamma_r \kappa_{sr}. \tag{54}$$

Finally, each elementary conflict redistributed by binary PCR6 preserves its total mass. Consequently, every node of a fixed Tree-PCR6 maps normalized BBAs with focal family $\mathcal{F}$ to normalized BBAs, and the final fused BBA is deterministic once the tree $\mathcal{T}$ is fixed.

Proof. Define

$$c_s = \frac{1}{S-1} \sum_{r \neq s} \left[1 - d_J(m_s, m_r)\right]. \tag{55}$$

The Jousselme distance is a metric on BBAs (Jousselme et al., 2001) and satisfies $0 \leq d_J \leq 1$ for normalized BBAs defined on the same frame. Hence $0 \leq c_s \leq 1$. Let $c_{\max} = \max_j c_j$. For

$$\gamma_s = \frac{c_s}{c_{\max} + \varepsilon}, \qquad \varepsilon > 0, \tag{56}$$

we have

$$0 \leq \gamma_s < 1. \tag{57}$$

The discounting rule can be written as

$$\boldsymbol{m}_s^{\gamma_s} = \gamma_s \boldsymbol{m}_s + (1-\gamma_s)\boldsymbol{m}_{\mathrm{V}}. \tag{58}$$

Therefore,

$$\boldsymbol{m}_s^{\gamma_s} - \boldsymbol{m}_{\mathrm{V}} = \gamma_s(\boldsymbol{m}_s - \boldsymbol{m}_{\mathrm{V}}). \tag{59}$$

Using the Jousselme metric (Jousselme et al., 2001),

$$\begin{aligned} d_J^2(m_s^{\gamma_s}, m_{\mathrm{V}}) &= \frac{1}{2}(\boldsymbol{m}_s^{\gamma_s} - \boldsymbol{m}_{\mathrm{V}})^{\top} D(\boldsymbol{m}_s^{\gamma_s} - \boldsymbol{m}_{\mathrm{V}}) \\ &= \gamma_s^2 d_J^2(m_s, m_{\mathrm{V}}). \end{aligned} \tag{60}$$

Taking square roots establishes Eq. equation 50.

For the focal family in Eq. equation 30, the pignistic probability is

$$p_{s,k} = m_s(\{\omega_k\}) + \frac{m_s(\Theta)}{K} \tag{61}$$

(Smets, 1990). After discounting,

$$\begin{aligned} p_{s,k}^{\gamma_s} &= \gamma_s m_s(\{\omega_k\}) + \frac{1-\gamma_s+\gamma_s m_s(\Theta)}{K} \\ &= \gamma_s p_{s,k} + \frac{1-\gamma_s}{K}. \end{aligned} \tag{62}$$

Stacking all components gives Eq. equation 51, while absolute homogeneity of norms yields Eq. equation 52.

Under Eq. equation 30, the only disjoint nonempty focal pairs are distinct singletons. Therefore,

$$\begin{aligned} \kappa_{sr}^{\gamma} &= \sum_{k\neq j} m_s^{\gamma_s}(\{\omega_k\}) m_r^{\gamma_r}(\{\omega_j\}) \\ &= \sum_{k\neq j} \gamma_s m_s(\{\omega_k\}) \gamma_r m_r(\{\omega_j\}) \\ &= \gamma_s \gamma_r \kappa_{sr}, \end{aligned} \tag{63}$$

which proves Eq. equation 54. Mass transferred to Θ introduces no additional conflict because $\Theta \cap A = A$ for every nonempty A.

Consider an elementary binary conflict between two disjoint focal elements with masses $a, b \geq 0$. In the two-source setting, PCR5 and PCR6 coincide (Martin & Osswald, 2006), and the conflicting mass ab is redistributed as

$$R_a = \frac{a^2 b}{a+b}, \qquad R_b = \frac{b^2 a}{a+b}, \qquad a+b>0. \tag{64}$$

Hence

$$R_a + R_b = \frac{a^2 b + b^2 a}{a+b} = ab, \tag{65}$$

with

$$0 \leq R_a \leq ab, \qquad 0 \leq R_b \leq ab. \tag{66}$$

When $a = b = 0$, all quantities are defined as zero.

Thus, every elementary conflict is redistributed without creating or discarding mass. Moreover, the focal family $\{\{\omega_k\}\}_{k=1}^{K} \cup \{\Theta\}$ is closed under all nonempty intersections arising in the conjunctive step, while conflicts between distinct singletons are redistributed back to the participating singletons. Therefore, every binary PCR6 node produces a normalized BBA with the same focal structure.

Applying this argument recursively and using induction shows that every intermediate node of Tree-PCR6, including the root, remains normalized. Because PCR6 is non-associative (Martin & Osswald, 2006), fixing $\mathcal{T}$ is necessary to make the recursive fusion operator deterministic. □

Corollary 1 (Suppression of unreliable evidence). As $\gamma_s \to 0$,

$$m_s^{\gamma_s} \to m_{\mathrm{V}}, \qquad \mathrm{BetP}(m_s^{\gamma_s}) \to \boldsymbol{u}, \tag{67}$$

and for every source r,

$$\kappa_{sr}^{\gamma} = \gamma_s \gamma_r \kappa_{sr} \to 0. \tag{68}$$

Thus, as a source becomes increasingly unreliable, it contributes progressively less committed class evidence and induces less pairwise conflict.

Limited dependence on the fusion tree. Let $\mathcal{P}$ denote the binary PCR6 operator. Although PCR6 is commutative, it is generally non-associative, and different source groupings may therefore produce different fused BBAs. Nevertheless, this dependence is controlled by the amount of conflict redistributed at the internal nodes. A binary PCR6 update can be written as

$$\mathcal{P}(m_i, m_j) = m_{\cap,ij} + R_{ij}, \qquad \|R_{ij}\|_1 = \kappa_{ij}, \tag{69}$$

where R_{ij} contains the redistributed partial conflicts and κ_{ij} is their total mass. Since conjunctive combination is associative, different fusion trees share the same limiting result when these conflict terms vanish. Under local Lipschitz continuity of the binary fusion operator, the discrepancy between two trees is therefore of order

$$\|m_{\mathcal{T}} - m_{\mathcal{T}'}\|_1 = \mathcal{O}\left(\sum_{v \in \mathcal{T} \cup \mathcal{T}'} \kappa_v\right), \tag{70}$$

where κ_v denotes the conflict encountered at an internal node.

In LFUF, reliability discounting further reduces the pairwise conflict induced by the original sources according to

$$\kappa_{sr}^{(\gamma)} = \gamma_s \gamma_r \kappa_{sr}. \tag{71}$$

Consequently, sources receiving low reliability contribute less to both the fused belief and the potential variation caused by alternative source groupings. The balanced tree also limits the number of successive binary updates. Thus, while Tree-PCR6 is not strictly invariant to the fusion tree, its dependence is expected to remain limited in the discounted low-conflict regime considered by LFUF. We use the same fixed balanced tree during training and inference to ensure a deterministic and reproducible fusion procedure.

Discussion. The two theorems characterize complementary robustness mechanisms in LFUF. Theorem 1 shows that choosing $1 < p < \alpha_{\mathrm{L}}$ places the recovery objective within the finite-moment regime of the underlying non-Gaussian stable source and yields a finite-step distributional stability guarantee under explicit rollout and transport anchoring conditions. Specifically,

$$W_p\left(\widehat{\mu}_1^{\mathcal{C}}, \nu_{\mathcal{C}}\right) \leq \frac{e^{L_s} - 1}{L_s}\left[(\ell_{\mathrm{SC}} + \epsilon^p)^{1/p} + \eta_{\mathrm{T}}\right]. \tag{72}$$

Theorem 2 establishes the exact relations

$$d_J(m_s^{\gamma_s}, m_{\mathrm{V}}) = \gamma_s d_J(m_s, m_{\mathrm{V}}), \qquad \kappa_{sr}^{\gamma} = \gamma_s \gamma_r \kappa_{sr}, \tag{73}$$

demonstrating that reliability discounting moves unreliable evidence toward vacuous belief while multiplicatively attenuating the conflict it induces. Binary PCR6 subsequently redistributes the remaining local conflict without changing its total mass.

These results do not imply that robust flow matching necessarily recovers the exact conditional field, that PCR6 necessarily improves classification accuracy, or that Tree-PCR6 is globally monotone with respect to every reliability coefficient. Such conclusions do not follow from the proposed operators alone. Rather, the analysis establishes the properties that can be rigorously guaranteed for LFUF: heavy-tail-compatible integrability, finite-step stability under explicit approximation conditions, exact reliability contraction, pairwise conflict attenuation, and local conservation of conflict mass.

B Effect of the Stability Parameter

We first examine the effect of the stability parameter α across different modality missing rates. The source scale is fixed at $\sigma = 1.0$, and the robust penalty $\rho_{p,\epsilon}$ uses $p = 1.05$ and $\epsilon = 10^{-3}$ throughout this analysis. All evaluated values of α satisfy $1 < p < \alpha$, while the remaining experimental settings are kept unchanged to isolate the effect of the source distribution.

As shown in Fig. 4, model performance varies non-monotonically with α. At a missing rate of 0.3, $\alpha = 1.8$ yields the highest accuracy and F1 score. At a missing rate of 0.5, the highest accuracy is achieved at approximately $\alpha = 1.7$, whereas the F1 score remains favorable over $\alpha = 1.5$–1.7.

The larger performance variation at this missing rate indicates greater sensitivity to the source distribution when the observed and missing information are relatively balanced. At the higher missing rate of 0.7, $\alpha = 1.7$–1.8 continues to provide strong and comparatively stable performance. By contrast, the Gaussian source, corresponding to $\alpha = 2.0$, does not attain the best performance at any evaluated missing rate.

These observations are consistent with the tail behavior of the α-stable distribution, for which the tail probability asymptotically satisfies $\Pr(|Z| > x) \propto x^{-\alpha}$ (Samorodnitsky & Taqqu, 1994; Nolan, 2020). Decreasing α produces heavier tails and increases the coverage of irregular or large latent deviations (Yoon et al., 2023; Pandey et al., 2025). However, along the flow path $\boldsymbol{X}_t = (1-t)\boldsymbol{Z}+t\boldsymbol{H}$ (Lipman et al., 2023), an excessively heavy-tailed source also increases the magnitude and variability of the target velocity $\boldsymbol{U} = \boldsymbol{H} - \boldsymbol{Z}$, making the recovery trajectory more difficult to estimate. This effect becomes particularly relevant at high missing rates, where fewer observed modalities are available to constrain strongly deviated trajectories. Meanwhile, the relatively small performance differences at a missing rate of 0.7 suggest that modifying the source distribution alone cannot fully compensate for substantial semantic information loss. At the other extreme, the light-tailed Gaussian source may provide insufficient coverage of the latent uncertainty introduced by missing modalities (Simsekli, 2017). Therefore, a moderately heavy-tailed source with α around 1.7–1.8 provides a favorable balance between latent coverage and recovery stability.

While α controls the tail behavior and latent-space coverage of the source distribution, p determines the growth rate of the robust penalty and hence the contribution of large velocity residuals during optimization. The two parameters therefore affect different aspects of recovery: α changes the sampled source distribution, whereas p changes the geometry of the regression objective. To verify that the benefit of the selected source distribution does not depend on a narrowly tuned penalty exponent, we complement the α analysis with a sensitivity study of p at fixed α.

Table 3: Sensitivity to the robust penalty exponent p on MELD at a modality missing rate of 0.5, with $\alpha = 1.8$.

p	1.1	1.2	1.3	1.4	1.5
Accuracy (%)	61.1	**61.2**	61.1	61.1	**61.2**
F1 (%)	57.7	**58.5**	58.4	58.1	58.3

Specifically, we fix $\alpha = 1.8$ and vary p from 1.1 to 1.5. All evaluated settings satisfy $1 < p < \alpha$. All evaluated settings satisfy $1 < p < \alpha$. As reported in Table 3, accuracy varies by less than 0.2 percentage points, while the maximum difference in F1 is 0.8 percentage points. Although $p = 1.2$ yields the highest F1 score, the main setting $p = 1.5$ achieves the same accuracy at the reported precision and differs by only 0.2 percentage points in F1. These small variations indicate that the model is not sensitive to the choice of p. Since $p = 1.5$ remains within the best-performing region and is consistently used in the main experiments, $(\alpha, p) = (1.8, 1.5)$ provides an appropriate and stable default setting.

C Comparison of Reliability Discounting and Evidence Combination Rules

To distinguish the contributions of reliability discounting and conflict redistribution, we compare different combinations of discounting strategies and fusion rules within the same framework. The multimodal encoder, missing-modality recovery module, evidence construction process, and all other training settings remain unchanged.

Throughout this comparison, LFUF employs the restricted focal family $\mathcal{F} = \{\{\omega_1\}, \ldots, \{\omega_K\}, \Theta\}$ and does not enumerate the full DSmT hyper-power set. PCR6 is used specifically as a local conflict-redistribution rule over these restricted BBAs.

As described in the method section, DST normalizes the conjunctive mass using the total conflict coefficient (Dempster, 1967; Shafer, 1976):

$$m_{\mathrm{DST}}(A) = \frac{m_{\cap}(A)}{1-\kappa}, \qquad \kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{74}$$

This operation applies the same normalization factor to all nonempty masses. PCR6 instead redistributes each partial conflict only among the hypotheses that generate it (Smarandache & Dezert,

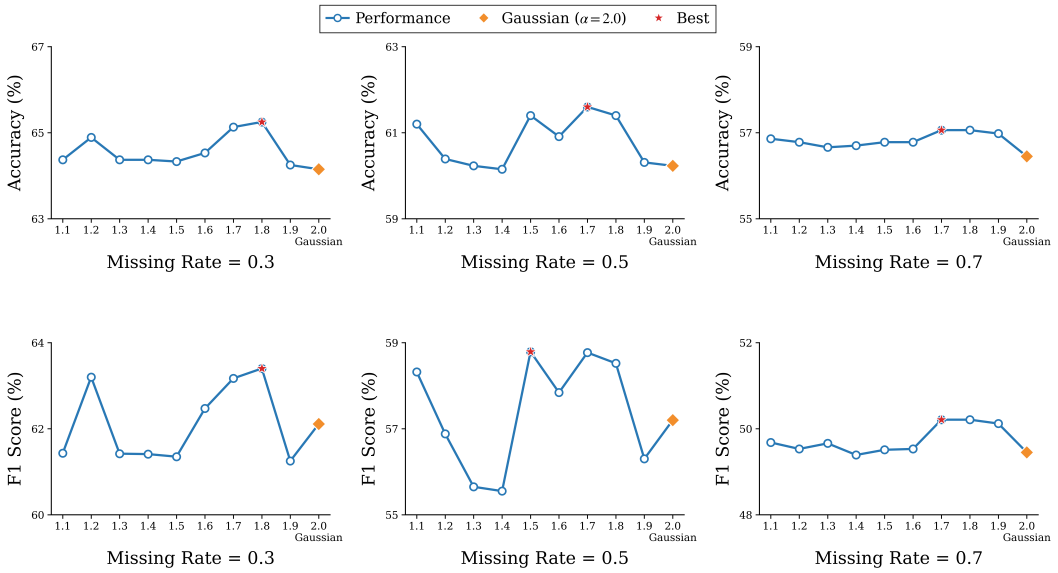


Figure 4: Accuracy and F1 score on MELD across different stability parameters α and modality missing rates. The robust penalty $\rho_{p,\epsilon}$ uses fixed parameters $p = 1.05$ and $\epsilon = 10^{-3}$. The case $\alpha = 2.0$ corresponds to a Gaussian source.

Table 4: Comparison of reliability discounting and evidence combination rules within the same framework on MELD. Each entry reports Accuracy/F1 (%) at the corresponding modality missing rate. The best result in each column is shown in bold.

Discounting	Fusion Rule	Missing Rate								Average
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	
None	DST	68.7/68.5	67.5/66.1	64.5/62.4	63.0/60.0	61.4/58.1	58.9/54.2	57.3/51.4	56.8/49.8	62.3/58.8
None	PCR6	69.2/68.3	67.5/66.2	65.8/64.2	63.9/60.8	61.8/58.8	60.3/56.2	58.4/52.3	56.9/49.8	63.0/59.6
Jousselme	Average	56.7/44.5	61.4/52.6	54.8/41.9	53.8/40.5	50.4/33.8	50.4/33.8	50.4/33.8	50.4/33.8	53.6/39.3
Jousselme	DST	69.3/68.3	67.5/66.3	65.5/64.0	63.0/60.8	62.0/58.8	59.8/56.4	57.8/**52.7**	56.4/48.2	62.7/59.4
Jousselme	**PCR6**	**69.6/69.1**	**68.3/67.3**	**66.6/64.8**	**64.8/62.8**	**62.3/59.1**	**61.2/58.3**	**58.8**/52.2	**57.2/50.4**	**63.6/60.5**

2006; Martin & Osswald, 2006). For a partial conflict ab, its redistributed components satisfy

$$\frac{a^2 b}{a+b} + \frac{b^2 a}{a+b} = ab. \tag{75}$$

Thus, PCR6 preserves the conflicting mass while retaining information about its origin.

As shown in Table 4, the complete model using Jousselme discounting and PCR6 (Jousselme et al., 2001; Mercier et al., 2008) obtains the highest average accuracy and F1 score. It also achieves the highest accuracy at every missing rate and the highest F1 score in all but one setting. Compared with undiscounted DST, the complete model improves the average accuracy and F1 by 1.3 and 1.7 percentage points, respectively. The gains are more evident at moderate missing rates, reaching 2.3 percentage points in accuracy and 4.1 percentage points in F1 at a missing rate of 0.5.

The controlled comparisons further separate the contributions of the two components. With PCR6 fixed, reliability discounting improves the average accuracy and F1 by 0.6 and 0.9 percentage points. With Jousselme discounting fixed, replacing DST with PCR6 improves them by 0.9 and 1.1 percentage points. The larger difference between the complete model and undiscounted DST therefore arises from the combined effect of reliability adjustment and local conflict redistribution.

The relatively moderate gain from discounting alone can be explained by the magnitude of the reliability coefficients. For a source with reliability γ_s, the distance between the discounted and original BBAs is

$$\left\| m_s^{(\gamma)} - m_s \right\|_1 = 2(1-\gamma_s)\big(1 - m_s(\Theta)\big). \tag{76}$$

When γ_s is close to one, discounting introduces only a small change. In our experiments, most sources agree on the dominant hypothesis and therefore receive relatively high reliability values. Discounting mainly suppresses a limited number of inconsistent sources while leaving the remaining evidence nearly unchanged.

Discounting also reduces the conflict passed to the fusion rule. For two sources, the effective conflict after discounting satisfies

$$\kappa^{(\gamma)} = \gamma_1 \gamma_2 \kappa. \tag{77}$$

Because unreliable singleton mass is transferred to Θ, it no longer contributes to pairwise conflict (Mercier et al., 2008). However, when both reliability coefficients are close to one, the reduction in κ remains limited, which explains the relatively small difference between discounted and undiscounted PCR6.

The difference between DST and PCR6 is similarly governed by the conflict magnitude (Zadeh, 1986; Yager, 1987; Smarandache & Dezert, 2006). When κ is small, their fused masses can be expressed as

$$m_{\mathrm{DST}}(A) = m_{\cap}(A) + \kappa m_{\cap}(A) + \mathcal{O}(\kappa^2), \tag{78}$$

and

$$m_{\mathrm{PCR6}}(A) = m_{\cap}(A) + r_A, \qquad \sum_A r_A = \kappa, \tag{79}$$

where r_A denotes the partial conflict redistributed to A. Both corrections are of order $\mathcal{O}(\kappa)$, so the two rules produce similar results under low conflict. Since Jousselme discounting further reduces the effective conflict, it also narrows the difference between discounted DST and discounted PCR6.

The variation across missing rates is consistent with this analysis. The coexistence of observed and recovered evidence can be approximated by $\rho(1-\rho)$, where ρ is the missing rate. This quantity is largest at intermediate missing rates, where observed and recovered sources are both influential and more likely to support different hypotheses. PCR6 is therefore more beneficial in this regime. At high missing rates, recovered sources tend to assign more mass to Θ, reducing singleton masses and their pairwise conflicts. The difference between DST and PCR6 consequently becomes smaller, while the limited semantic information becomes the main performance constraint.

Arithmetic averaging performs substantially worse because

$$m_{\mathrm{avg}} = \frac{1}{S} \sum_{s=1}^{S} m_s^{(\gamma)} \tag{80}$$

assigns the same coefficient to every source without considering agreement or conflict. Reliable evidence may therefore be diluted by uncertain sources, demonstrating the need to model evidence interactions explicitly in the seven-source setting.

Finally, changes in fused masses do not always alter the predicted class. Let $\Delta_p = p_{(1)} - p_{(2)}$ denote the margin between the two largest class probabilities. If a change caused by an alternative fusion rule satisfies

$$\|\delta p\|_\infty < \frac{\Delta_p}{2}, \tag{81}$$

the predicted class remains unchanged. Reliability discounting and PCR6 may therefore modify confidence, ignorance mass, and conflict allocation while producing only limited changes in accuracy and F1.

D Analysis of Shortcut Self-Consistency

We further analyze the effect of Shortcut self-consistency (SC) on missing-modality recovery. The modality missing rate is fixed at 0.3, and each model is evaluated using different numbers of function evaluations (N_{FE}). Standard flow matching learns an instantaneous velocity field, while the resulting inference trajectory can still depend on the integration step size and accumulated discretization errors (Lipman et al., 2023; Tong et al., 2024). Shortcut models additionally learn finite-step transport maps to improve consistency across different inference discretizations (Frans et al., 2025).

Let $\widehat{\Phi}_{t,t+h}$ denote the learned transport map over an interval of length h. Its finite-step consistency defect can be written as

$$D_h(\boldsymbol{x}, t) = \left\| \widehat{\Phi}_{t,t+2h}(\boldsymbol{x}) - \widehat{\Phi}_{t+h,t+2h}\left(\widehat{\Phi}_{t,t+h}(\boldsymbol{x})\right) \right\|_2 . \tag{82}$$

This quantity measures the discrepancy between coarse and refined updates that arrive at the same time point. If the finite-step maps satisfy a Lipschitz condition, the terminal discrepancy between the corresponding rollouts is controlled by the accumulated consistency defects:

$$\left\| \widehat{\boldsymbol{X}}_1^{\text{coarse}} - \widehat{\boldsymbol{X}}_1^{\text{fine}} \right\| \leq \sum_{j=0}^{N-1} K^{N-1-j} D_{h,j}, \tag{83}$$

where K denotes the stability factor of the finite-step map. Therefore, reducing the local consistency defect limits the divergence between recovery trajectories produced using different inference step sizes.

Since SC primarily targets consistency across inference discretizations, we evaluate its effect through the variation in L_{rec} across different inference budgets. We define this variation as

$$\Delta_{\text{rec}} = \max_{n \in \mathcal{N}} \mathcal{L}_{\text{rec}}^{(n)} - \min_{n \in \mathcal{N}} \mathcal{L}_{\text{rec}}^{(n)}, \tag{84}$$

where $\mathcal{N}$ is the evaluated N_{FE} set. A smaller Δ_{rec} indicates that the recovered representations are less sensitive to the choice of inference step size.

Table 5: Cross-step L_{rec} variation and accuracy at $N_{\text{FE}} = 8$ under a missing rate of 0.3.

Method	IEMOCAP4		MELD	
	Δ_{rec}	Acc. (%)	Δ_{rec}	Acc. (%)
SC	2.4×10^{-5}	85.0	2.6×10^{-5}	64.8
No-SC	2.1×10^{-2}	84.5	2.1×10^{-2}	62.4

As shown in Table 5, the maximum L_{rec} variation of Shortcut remains on the order of 10^{-5} on both datasets, whereas that of No-SC reaches the order of 10^{-2}. This result indicates that the No-SC recovery is substantially more sensitive to the selected inference step size. SC reduces this sensitivity and produces more consistent recovery trajectories across different N_{FE} settings.

Under the same inference budget, Shortcut also maintains competitive classification accuracy on both IEMOCAP4 and MELD. These results identify the primary contribution of SC as improving finite-step recovery stability.