FROM UNIQUE TO UNIFIED: ATTRIBUTE-CENTRIC SELF-SUPERVISED REPRESENTATION LEARNING FOR MIXED-TYPE TABULAR DATA CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Mixed-type tabular data are common in scientific and real-world applications, yet learning reliable representations remains challenging because their attributes encode different relational semantics. Current self-supervised paradigms commonly derive supervision from inter-instance similarity, but such relations can be insufficient or unstable for heterogeneous tabular data. Moreover, numerical, nominal, and ordinal attributes encode fundamentally different notions of similarity. We therefore exploit inter-attribute dependencies as a complementary source of self supervision and follow a preserve-before-unify principle, in which type-specific semantics are first retained and subsequently integrated into a shared representation. Based on this principle, we propose **F**aithful **U**nification through **S**emantics-Preserving **E**ncoding (FUSE), an attribute-centric self-supervised framework for faithfully unifying mixed attributes. Its **T**ype-Faithful **A**ttribute **F**ormation (TAF) module aligns perturbation and recovery objectives with the relational properties of each attribute type, preserving their unique semantics before fusion. The **C**ross-Attribute **R**elational **U**nification (CRU) module further captures within-sample cross-attribute dependencies through target-exclusive contextual prediction, providing relational supervision for a unified representation. Comprehensive experiments across diverse mixed-type benchmarks demonstrate the efficiency and superiority of FUSE. Code is available at FUSE.

1 INTRODUCTION

Mixed-type tabular data arise commonly in domains such as healthcare, finance, and social sciences, where numerical measurements, nominal categories, and ordinal assessments coexist within the same dataset Hunt & Jorgensen (2011). Clustering is a central tool for exploratory analysis because it reveals latent organization without requiring annotations Jain et al. (1999). However, the mixed attributes express distinct relational semantics. Numerical values quantify continuous variation, nominal values distinguish unordered categories, and ordinal values convey ordered but discrete levels. Consequently, an effective data representation that retains the inherently specific structure for each type while capturing how the attributes jointly characterize each sample is urgently needed.

Clustering mixed-type data inherently requires bridging attributes with distinct relational semantics, a challenge exacerbated by the lack of natural geometry in categorical attributes. Conventional paradigms address this by engineering comparable metrics, moving from simple categorical mismatches paired with numerical differences Gower (1971); Huang (1997; 1998) to more nuanced conditional statistics Ienco et al. (2009), association-based dissimilarities Le & Ho (2005); Ahmad & Dey (2007), and unified similarity metrics Cheung & Jia (2013); Zhang et al. (2020). Expanding on these associations, graph-based methods organize attribute couplings into structural object graphs to learn latent representations Chen et al. (2024). Despite improving heterogeneous comparisons, these approaches share a common characteristic in that they explicitly predefine the form of attribute relationships. To move beyond static statistical or geometric assumptions, recent self-supervised learning paradigms offer a compelling alternative by implicitly learning rich, data-driven representations directly from the raw heterogeneous inputs.

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.

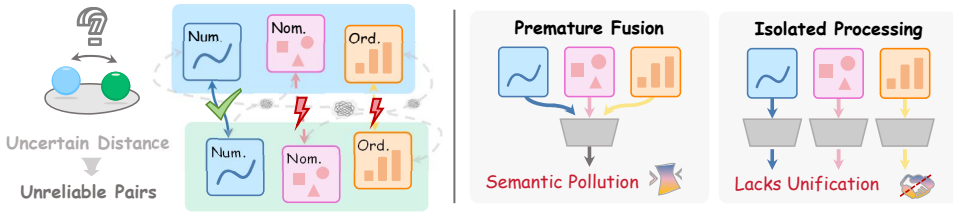

Figure 1: Motivation of FUSE. Left: Heterogeneous attribute semantics complicate the construction of reliable sample-level relations. Right: Two architectural tendencies for mixed-type representation learning. Premature Fusion may cause semantic interference by applying a common objective before type-specific semantics are adequately preserved, whereas Isolated Processing may limit cross-attribute interactions after type-specific encoding.

Within this paradigm, training signals are derived from reconstructing corrupted attributes or organizing representations through augmented views, encouraging information retention and representation robustness Yoon et al. (2020); Bahri et al. (2022). However, for mixed-type data, constructing and interpreting sample-level relations remains challenging because attributes express relational evidence under disparate semantics, as illustrated in the left panel of Figure 1. Existing tabular contrastive methods like SubTab Ucar et al. (2021) and SCARF Bahri et al. (2022) rely on predefined instance-level similarities or noise-induced views, which become unreliable when heterogeneous attribute semantics have not been unified. Nevertheless, mixed-type observations provide a complementary information source, as interactions among attributes describing the same sample enrich its representation with contextual cues. This within-sample conditional context offers an attribute-level supervisory basis for learning how heterogeneous attributes jointly characterize observations.

Two strategies for using such conditional information in representation learning are early fusion Yoon et al. (2020); Somepalli et al. (2021); Ucar et al. (2021) and type-specific processing Zhang et al. (2022); Tokar & Sanner (2025); Luo et al. (2024), as illustrated in the right panel of Figure 1. Early fusion facilitates information exchange across attributes, but applying a common learning objective before their distinct semantics are preserved may introduce semantic interference. Type-specific processing helps retain the intrinsic structure of each attribute type, although combining the representations may provide limited integration of conditional dependencies across attributes. These observations motivate preserving type-specific relational semantics before incorporating cross-attribute information into a unified representation.

To this end, FUSE (**F**aithful **U**nification through **S**emantics-Preserving **E**ncoding) is proposed as an attribute-centric self-supervised framework that structures mixed-type representation learning according to a *preserve-before-unify* principle. To encourage semantic distinction and prevent manifold distortion, Type-Faithful Attribute Formation (TAF) first crystallizes type-specific semantics through perturbation and recovery objectives tailored to the distinct geometries of numerical, nominal, and ordinal attributes. Building upon these preserved semantics, Cross-Attribute Relational Unification (CRU) eliminates identity shortcut learning by fully withholding each target attribute, forcing the latent space to capture genuine cross-attribute conditional synergy without requiring a predefined sample metric. Crucially, by isolating representation formation from downstream partition optimization, FUSE prevents early latent spaces from collapsing into provisional clustering bias, yielding a universally applicable semantic foundation that preserves intrinsic data structures.

The main contributions are summarized as follows:

- Establishes a novel attribute-centric self-supervised representation learning paradigm for mixed-type data. Within-sample conditional dependencies are elevated to a primary attribute-level learning signal, providing the relational bridge from type-specific semantics to downstream sample organization.

- Formulates Type-Faithful Attribute Formation (TAF), which uses type-matched perturbation-recovery schemes to mitigate semantic interference and establish robust and structure-preserved embeddings for the subsequent unified representation learning.

- Develops Cross-Attribute Relational Unification (CRU) to expose conditional dependencies among attributes through target-exclusive contextual prediction, consolidating attribute-wise predictability into a single comprehensive unified relational structure.

2 Related Work

Mixed-Type Data Clustering. Classical methods construct comparable dissimilarities across heterogeneous attributes, including K-Prototypes and its variants Huang (1997; 1998), frequency-based self-organizing maps Del Coso et al. (2015), context-based distances Ienco et al. (2009), association-based value dissimilarities Le & Ho (2005); Ahmad & Dey (2007), unified similarity metrics Cheung & Jia (2013); Zhang et al. (2025a), and adaptive partitioning schemes Zhang et al. (2024; 2025b). Representation-based methods instead learn a latent space before clustering. Heterogeneous coupling learning models value interdependencies Zhu et al. (2022), Het2Hom maps different attribute types into homogeneous concept spaces Zhang et al. (2022), and graph-based approaches encode structured relations among heterogeneous observations Zhang & Cheung (2023); Chen et al. (2024; 2026). These methods improve mixed-type comparison, but their relational structures are commonly specified through designed similarities, couplings, or graphs. FUSE instead derives its representation-learning signals directly from unlabeled attributes while strictly preserving their unique type-specific semantics.

Self-Supervised Representation Learning. Self-supervised methods for structured data derive supervision by reconstructing corrupted information, predicting contextual targets, or aligning related views. Reconstruction- and prediction-based methods range from attribute and sample recovery in VIME and SubTab Yoon et al. (2020); Ucar et al. (2021) to latent-target prediction in T-JEPA Thimonier et al. (2025). To handle the heterogeneity of prediction targets, Binning-based pretext learning converts numerical reconstruction into bin-index prediction, placing numerical and categorical attributes under a common categorical target space Lee et al. (2024). Contrastive methods construct correspondence between corrupted versions of a sample Bahri et al. (2022); Somepalli et al. (2021), while ICE-T contrasts an attribute with the complementary attributes of the same sample Tokar & Sanner (2025). Multi-view approaches further exploit instance consistency, cross-view recovery, or cluster-level alignment Luo et al. (2024); Su et al. (2025); Dong et al. (2026). FUSE makes the ordering of semantic formation and cross-attribute interaction an explicit architectural principle, forming type-faithful representations before modeling their dependencies in a shared space.

3 Method

3.1 Problem Formulation and Overview

Consider a mixed-type dataset with N samples whose attributes may be numerical, nominal, or ordinal. For each sample i, let

$$x_i = \left(x_i^{(n)}, x_i^{(c)}, x_i^{(o)} \right), \tag{1}$$

where $x_i^{(n)} \in \mathbb{R}^{p_n}$, while $x_i^{(c)}$ and $x_i^{(o)}$ contain p_c nominal and p_o ordinal attributes, respectively. A nominal attribute j has C_j categorical values, and an ordinal attribute j has L_j ordered levels.

FUSE follows a preserve-before-unify design, as illustrated in Figure 2. TAF forms type-faithful representations through type-matched perturbation and recovery, as described in Section 3.2. CRU then uses target-exclusive contextual prediction to supervise the unified encoder, as detailed in Section 3.3. The training objective and downstream clustering procedure are presented in Section 3.4.

3.2 Type-Faithful Attribute Formation

Different self-supervised pretexts can expose different aspects of tabular data, while a given corruption operation may carry different meanings across attribute types. For example, Gaussian perturbation expresses numerical variation but has no natural interpretation for nominal values, whose categories have no inherent numerical proximity. TAF therefore matches each attribute type with a compatible input representation, corruption process, and recovery objective, aligning the resulting supervision with the relational semantics of the attribute.

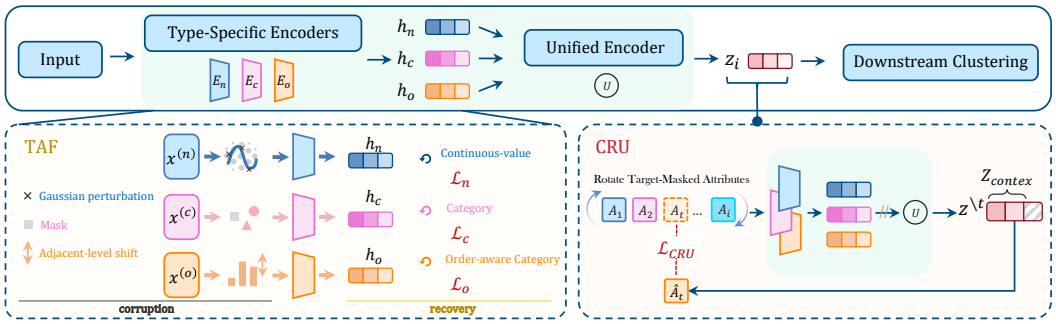

Figure 2: Overview of FUSE. Input attributes are first processed by type-specific encoders to form $h^{(n)}$, $h^{(c)}$, and $h^{(o)}$, which are jointly processed by the unified encoder U to produce the representation z_i. Specifically, TAF applies type-matched perturbation and recovery to the three attribute types, preserving their relational semantics through $\mathcal{L}_n$, $\mathcal{L}_c$, and $\mathcal{L}_o$. Building on these type-faithful encodings, CRU withholds one target attribute and predicts it from the remaining attributes through U. This target-exclusive contextual prediction captures how attributes jointly characterize each sample and provides relational supervision for the unified representation used in downstream clustering.

Numerical Attributes. For numerical attributes, local perturbations provide a natural recovery signal. Let $m^{(n)}$ indicate the values retained after random masking. Masked values are filled with zero, while retained values receive Gaussian noise; the visibility indicator distinguishes masked entries from observed zeros:

$$\tilde{x}^{(n)} = m^{(n)} \odot x^{(n)} + m^{(n)} \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma_n^2 I). \tag{2}$$

The numerical encoder therefore receives both the corrupted values and their indicators,

$$h^{(n)} = E_n([\tilde{x}^{(n)}; m^{(n)}]), \tag{3}$$

and a regression head recovers the hidden continuous values. The resulting mean squared error supervises continuous-value recovery:

$$\mathcal{L}_n = \frac{\sum_{i,j}(1 - m_{ij}^{(n)}) \left(\hat{x}_{ij}^{(n)} - x_{ij}^{(n)}\right)^2}{\sum_{i,j}(1 - m_{ij}^{(n)}) + \epsilon_0}. \tag{4}$$

Nominal Attributes. Nominal attributes encode category identity without an intrinsic ordering. Each nominal attribute j has an individual embedding table for its C_j categories and one additional index for hidden values. This avoids introducing an artificial scalar order between categories.

The category embeddings can be initialized from unlabeled category-context statistics. For each category of attribute j, the initialization summarizes the empirical distributions of the remaining nominal attributes, the ordinal levels, and discretized numerical values. Distances between these summaries are projected into the embedding space. The initialization provides a context-informed starting point and remains trainable throughout learning.

The nominal input is further augmented with a statistical profile $s_{ij}^{(c)}$ for the category observed in sample i. The profile contains its empirical frequency and the category-wise means and standard deviations of the numerical attributes. After projection, it is concatenated with the category embedding:

$$v_{ij}^{(c)} = \mathrm{Emb}_j(\tilde{x}_{ij}^{(c)}) \,\|\, P_c(s_{ij}^{(c)}), \qquad h_i^{(c)} = E_c\left(\|_{j=1}^{p_c} v_{ij}^{(c)}\right). \tag{5}$$

When a nominal value is hidden, its additional embedding index is used and its profile is removed. An attribute-specific categorical head then predicts the clean category. Let $\Omega_j^{(c)}$ denote the samples in which nominal attribute j is hidden. Its recovery objective is

$$\mathcal{L}_c = \frac{1}{p_c} \sum_{j=1}^{p_c} \frac{1}{|\Omega_j^{(c)}|} \sum_{i \in \Omega_j^{(c)}} \mathrm{CE}\left(D_{c,j}(h_i^{(c)}), x_{ij}^{(c)}\right). \tag{6}$$

Both category-context initialization and profile augmentation use no class labels. They support nominal formation without replacing the cross-attribute supervision introduced by CRU.

Ordinal Attributes. Ordinal attributes contain discrete identities together with an ordering among their levels. Treating them as nominal would discard this order, while treating them as continuous would assume unrestricted metric intervals. TAF retains both properties through a level embedding and an explicit normalized rank. Randomly selected values are hidden, and retained values may shift to an adjacent level while remaining within the valid range. For ordinal attribute j, its input token is

$$v_{ij}^{(o)} = \mathrm{Emb}_j(\tilde{x}_{ij}^{(o)}) \parallel \bar{x}_{ij}^{(o)} \parallel m_{ij}^{(o)}, \qquad h_i^{(o)} = E_o\left(\|_{j=1}^{p_o} v_{ij}^{(o)}\right), \tag{7}$$

where $\bar{x}_{ij}^{(o)} = \tilde{x}_{ij}^{(o)}/(L_j - 1)$ for a visible value and zero otherwise. A hidden value uses an additional index outside of the valid ordered levels.

Recovery is formulated through cumulative thresholds Frank & Hall (2003); Niu et al. (2016). The target at threshold k is

$$y_{ijk}^{\mathrm{cum}} = \mathbb{I}[x_{ij}^{(o)} > k], \qquad k = 0, \ldots, L_j - 2. \tag{8}$$

The cumulative predictions encode the ordering between levels. Their summed probabilities also define the expected rank

$$\hat{r}_{ij} = \sum_{k=0}^{L_j-2} \sigma\left(\hat{y}_{ijk}^{\mathrm{cum}}\right). \tag{9}$$

The recovery objective combines cumulative binary cross-entropy with the normalized distance between the expected and clean ranks:

$$\mathcal{L}_o = \underset{(i,j):m_{ij}^{(o)}=0}{\mathrm{Avg}} \left[\mathrm{BCE}\left(\hat{\mathbf{y}}_{ij}^{\mathrm{cum}}, \mathbf{y}_{ij}^{\mathrm{cum}}\right) + \lambda_o \frac{|\hat{r}_{ij} - x_{ij}^{(o)}|}{L_j - 1}\right]. \tag{10}$$

Unification. The recovery heads operate directly on their respective branch encodings. The three type-specific encodings are then concatenated and mapped by the unified encoder:

$$z = U([h^{(n)}; h^{(c)}; h^{(o)}]). \tag{11}$$

For a single corrupted input, the type-faithful formation objective is

$$\mathcal{L}_{\mathrm{TAF}} = \mathcal{L}_n + \mathcal{L}_c + \mathcal{L}_o. \tag{12}$$

3.3 Cross-Attribute Relational Unification

TAF faithfully preserves the unique semantics of each attribute type, but its recovery heads supervise the branch encodings separately. These objectives do not require the unified representation to retain how different attributes jointly characterize a sample. Cross-Attribute Relational Unification (CRU) addresses this gap by predicting one attribute from the context provided by the others. Success across heterogeneous targets requires the unified representation to integrate information that is distributed across the type-specific branches.

CRU differs from the random value masking used by TAF. TAF independently hides individual values and applies recovery supervision to the corresponding type-specific branch. CRU instead creates a dedicated prediction pass for a target attribute t and withholds that attribute for every sample in the batch. The distinction concerns both the masking scope and the representation being supervised. The CRU prediction head receives only the unified representation.

For each batch $\mathcal{B}$, a target set $\mathcal{T}_{\mathcal{B}}$ is sampled from all attributes. Given a corrupted input $\tilde{x}_i$, the withholding operator $\mathcal{W}_t$ removes the observed target value and its associated side information for every $i \in \mathcal{B}$:

$$\tilde{x}_i^{\setminus t} = \mathcal{W}_t(\tilde{x}_i), z_i^{\setminus t} \quad = U\left([h_i^{(n),\setminus t}; h_i^{(c),\setminus t}; h_i^{(o),\setminus t}]\right), \hat{y}_{i,t} = D_t(z_i^{\setminus t}). \tag{13}$$

Here, each $h_i^{(r),\setminus t}$ is produced by the corresponding TAF encoder, and an unavailable attribute type is omitted. The input retains only a type-compatible representation of the target's absence. The target $x_{i,t}$ is used to compute the prediction loss. It cannot enter $z_i^{\setminus t}$ through its value representation or, for a nominal target, through its statistical profile.

Each attribute has its own prediction head, and its target follows the semantics established by TAF. A numerical target is assigned to an empirical quantile bin, which provides a coarse prediction task while numerical recovery retains continuous variation. A nominal target is predicted as an unordered category. An ordinal target is represented by the cumulative threshold vector defined in the previous section. The corresponding loss is

$$\ell_t(\hat{y}_{i,t}, x_{i,t}) = \begin{cases} \mathrm{CE}(\hat{y}_{i,t}, b_t(x_{i,t})), & t \text{ is numerical}, \\ \mathrm{CE}(\hat{y}_{i,t}, x_{i,t}), & t \text{ is nominal}, \\ \mathrm{BCE}(\hat{y}_{i,t}, \mathbf{y}_{i,t}^{\mathrm{cum}}), & t \text{ is ordinal}, \end{cases} \tag{14}$$

where b_t denotes the quantile-bin assignment for numerical attribute t. To enforce genuine cross-attribute conditional synergy within the unified space, the CRU objective minimizes the overall contextual prediction error across the selected targets and samples:

$$\mathcal{L}_{\mathrm{CRU}} = \frac{1}{|\mathcal{T}_{\mathcal{B}}|} \sum_{t \in \mathcal{T}_{\mathcal{B}}} \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \ell_t(\hat{y}_{i,t}, x_{i,t}). \tag{15}$$

To strictly keep the type-specific formation and relational unification at distinct representation levels, the branch encodings are detached along the CRU path in the default optimization. Thus, $\mathcal{L}_{\mathrm{CRU}}$ updates the unified encoder and the target-specific heads, while $\mathcal{L}_{\mathrm{TAF}}$ directly trains the three independent type-specific branches. By dynamically resampling at most M targets across successive batches, CRU ensures comprehensive attribute coverage over the course of training while bounding the per-batch computational overhead to $O(M)$ target-specific passes.

3.4 Model Training and Clustering

TAF and CRU provide the two distinct primary learning signals of FUSE. The former forms type-specific encodings, while the latter supervises their integration in the unified space. During optimization, each sample is presented through two input versions with different corruption strengths. The TAF recovery is evaluated on both versions, and CRU operates on their target-withheld inputs.

The paired inputs also support a small auxiliary regularizer. Let $z_i^{(w)}$ and $z_i^{(s)}$ denote their unified representations, and let $q_i^{(v)} = P(z_i^{(v)})$ be the output of a training-only projection head for $v \in \{w, s\}$. The regularizer encourages the two projections of a sample to remain close. A variance term prevents individual dimensions from becoming inactive, and a covariance term reduces redundancy between dimensions:

$$\mathcal{L}_{\mathrm{reg}} = \alpha \mathcal{L}_{\mathrm{inv}} + \beta \mathcal{L}_{\mathrm{var}} + \gamma \mathcal{L}_{\mathrm{cov}}. \tag{16}$$

Here, $\mathcal{L}_{\mathrm{inv}}$ is the mean squared distance between $q^{(w)}$ and $q^{(s)}$. The variance term penalizes a dimension whose standard deviation falls below a target value, and a covariance term penalizes squared off-diagonal covariance, following the variance-invariance-covariance regularization paradigm established in non-contrastive self-supervised learning Bardes et al. (2022); Zbontar et al. (2021). This standard regularizer supports training stability. It does not define either type-faithful formation or relational unification. The projection head P is discarded after training. Thus, the main training objective is

$$\mathcal{L} = \lambda_{\mathrm{TAF}} \mathcal{L}_{\mathrm{TAF}} + \lambda_{\mathrm{CRU}} \mathcal{L}_{\mathrm{CRU}} + \lambda_{\mathrm{reg}} \mathcal{L}_{\mathrm{reg}}. \tag{17}$$

For downstream clustering evaluation, K-means is applied to the unified representations obtained after FUSE training.

Complexity. Let N, p, B, T, and M denote the numbers of samples and attributes, the batch size, the training epochs, and the CRU target budget, respectively. For network width d and maximum target-output size $C_{\max}$, FUSE requires $\mathcal{O}\big(TN[(M+1)(pd+d^2) + MdC_{\max}]\big)$ training time and $\mathcal{O}(|\Theta| + B(M+1)(p+d+C_{\max}))$ peak memory. Since M is bounded, fixed M, d, and $C_{\max}$ yield linear sample and attribute scaling and peak batch memory independent of N. The computational cost remains within the range of the compared deep self-supervised methods.

4 EXPERIMENTS

This section evaluates FUSE against representative baselines on mixed-type benchmarks. The evaluation covers overall clustering performance (Section 4.2), ablation studies (Section 4.3), convergence analysis (Section 4.4), and robustness and scalability (Section 4.5). Additional method details, experimental settings, and analyses are provided in the appendix.

4.1 EXPERIMENTAL SETUP

Datasets. Experiments evaluate FUSE across 11 benchmark datasets sourced from the UCI Machine Learning Repository Dua & Graff (2019) and OpenML Vanschoren et al. (2014). These datasets encompass highly diverse attribute compositions, ranging from purely nominal attributes to complex mixed-type structures with varying proportions of numerical, nominal, and ordinal features. The benchmark scale varies significantly, from small-scale datasets with roughly 100 samples to large-scale real-world scenarios containing over 34,000 observations, while the number of underlying clusters spans from 2 to 7. Detailed dataset statistics are provided in Appendix C.1.

Baselines. The baselines encompass four complementary groups: traditional mixed-type clustering (OH-KM and K-Prototypes/K-Modes/K-means Huang (1997)), relation-aware heterogeneous modeling (HARR Zhang et al. (2025a) and HyReaL Chen et al. (2026)), deep multi-view clustering (SCM Luo et al. (2024), MGBCC Su et al. (2025), and RAC Dong et al. (2026)), and tabular self-supervised learning (SubTab Ucar et al. (2021), ICE-T Tokar & Sanner (2025), and T-JEPA Thimonier et al. (2025)). Notably, deep multi-view methods are incorporated to evaluate whether view-level integration can substitute type-specific representation formation, with inputs structured as sample-aligned type-based views.

Preprocessing and evaluation. Numerical attributes are standardized, nominal attributes are mapped into discrete identifiers, and ordinal attributes are encoded by their ordered levels. Method-specific input transformations, including one-hot encoding and type-based view construction, follow input requirements of each baseline. For representation-based methods, K-means is applied with the ground-truth cluster count and ten initializations. Clustering quality is measured by Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI).

Implementation details. FUSE is trained for 20 epochs with a batch size of 128 using the Adam optimizer with a weight decay of 10^{-4}. The learning rate is selected from $\{5\times10^{-4}, 10^{-3}\}$. Each type-specific branch contains two hidden layers with 512 and 256 units. Their outputs are integrated by a 256-unit fusion layer to produce a 128-dimensional unified representation. Results are reported as the mean and standard deviation over five runs with different random seeds. Additional parameter sensitivity analyses are provided in Appendix C.4.1.

4.2 OVERALL CLUSTERING PERFORMANCE

To assess FUSE across different dataset scales and attribute compositions, we compare it with ten representative baselines on eleven mixed-type datasets using ACC, NMI, and ARI. Table 1 reports the main comparisons on nine datasets, while the results for AC and CMC are provided in the appendix. All results follow the unified evaluation protocol and are reported as the mean and population standard deviation over runs.

Across the nine datasets in Table 1, FUSE achieves the best result in 24 of 27 dataset–metric comparisons. It ranks first in ACC and ARI on all nine datasets and in NMI on six. The gains span different attribute compositions. FUSE ranks first on BC and DG, both of which contain ordinal attributes, and also performs strongly on ZOO, a nominal-only dataset. This pattern is consistent with the role of TAF in matching recovery signals to attribute semantics, while the gains on larger mixed-type datasets such as DC, ADT, and NOM are compatible with CRU's use of cross-attribute dependencies. The remaining wins of ICE-T and OH-KM suggest that complementary-attribute contrastive learning and conventional metric-based clustering can still be advantageous for particular dataset–metric combinations. These results show that FUSE performs consistently across datasets with different attribute compositions and scales, consistent with its preserve-before-unify design.

	Method	ZOO	BC	HF	DG	ANL	DRUG	DC	ADT	NOM	Avg.Rank
ACC	OH-KM	0.751 ± 0.06	0.736 ± 0.00	0.756 ± 0.01	0.727 ± 0.01	0.450 ± 0.00	0.251 ± 0.02	0.685 ± 0.00	0.684 ± 0.00	0.753 ± 0.00	7.375
	K-Pro. (PAKDD'97)	0.673 ± 0.00	0.652 ± 0.04	0.719 ± 0.04	0.735 ± 0.12	0.377 ± 0.04	0.220 ± 0.02	0.693 ± 0.00	0.675 ± 0.04	0.753 ± 0.00	9.125
	HARR (ESWA 2025)	0.713 ± 0.04	0.721 ± 0.02	0.581 ± 0.02	0.678 ± 0.17	0.402 ± 0.03	0.264 ± 0.05	0.558 ± 0.07	0.668 ± 0.02	0.769 ± 0.05	8.125
	HyReaL (DASFAA'26)	0.808 ± 0.01	0.690 ± 0.00	0.668 ± 0.05	0.659 ± 0.06	0.522 ± 0.05	0.392 ± 0.01	–	–	–	6.250
	SCM (IJCAI'24)	0.861 ± 0.07	0.706 ± 0.03	0.587 ± 0.01	0.900 ± 0.03	0.461 ± 0.03	0.425 ± 0.06	0.574 ± 0.02	0.708 ± 0.05	0.667 ± 0.02	4.750
	MGBCC (AAAI'25)	0.871 ± 0.03	0.747 ± 0.00	0.644 ± 0.02	0.937 ± 0.00	0.393 ± 0.02	0.429 ± 0.00	0.514 ± 0.00	0.668 ± 0.05	0.663 ± 0.09	3.875
	RAC (AAAI'26)	0.885 ± 0.01	0.731 ± 0.02	0.715 ± 0.02	0.448 ± 0.04	0.453 ± 0.02	0.360 ± 0.01	0.529 ± 0.02	0.720 ± 0.00	0.718 ± 0.08	5.625
	SubTab (NeurIPS'21)	0.752 ± 0.08	0.706 ± 0.05	0.769 ± 0.07	0.932 ± 0.01	0.470 ± 0.05	0.284 ± 0.01	0.578 ± 0.03	0.711 ± 0.00	0.759 ± 0.03	5.750
	ICE-T (AAAI'25)	0.893 ± 0.01	0.561 ± 0.09	0.549 ± 0.01	0.762 ± 0.01	0.507 ± 0.08	0.339 ± 0.02	0.767 ± 0.00	0.658 ± 0.01	0.783 ± 0.01	5.875
	T-JEPA (ICLR'25)	0.835 ± 0.08	0.514 ± 0.00	0.557 ± 0.03	0.738 ± 0.00	0.505 ± 0.06	0.304 ± 0.02	0.613 ± 0.03	0.716 ± 0.00	0.750 ± 0.02	7.875
	FUSE (Ours)	**0.904** ± 0.02	**0.759** ± 0.01	**0.781** ± 0.02	**0.977** ± 0.00	**0.664** ± 0.02	**0.442** ± 0.01	**0.801** ± 0.00	**0.739** ± 0.00	**0.804** ± 0.01	**1.375**
NMI	OH-KM	0.825 ± 0.02	0.090 ± 0.00	0.184 ± 0.01	0.868 ± 0.01	0.143 ± 0.00	0.087 ± 0.00	0.001 ± 0.00	**0.176** ± 0.00	0.127 ± 0.00	6.375
	K-Pro. (PAKDD'97)	0.764 ± 0.00	0.066 ± 0.02	0.141 ± 0.06	0.841 ± 0.06	0.120 ± 0.03	0.060 ± 0.00	0.001 ± 0.00	0.126 ± 0.01	0.119 ± 0.02	8.375
	HARR (ESWA 2025)	0.833 ± 0.03	0.036 ± 0.04	0.018 ± 0.00	0.746 ± 0.11	0.117 ± 0.01	0.058 ± 0.00	0.000 ± 0.00	0.128 ± 0.01	0.186 ± 0.09	8.375
	HyReaL (DASFAA'26)	0.770 ± 0.03	0.093 ± 0.00	0.050 ± 0.03	0.588 ± 0.07	0.038 ± 0.04	0.041 ± 0.01	–	–	–	8.125
	SCM (IJCAI'24)	0.875 ± 0.03	0.078 ± 0.00	0.008 ± 0.01	0.863 ± 0.04	0.308 ± 0.05	0.104 ± 0.01	0.012 ± 0.01	0.060 ± 0.04	0.047 ± 0.02	4.875
	MGBCC (AAAI'25)	0.866 ± 0.05	0.112 ± 0.01	0.038 ± 0.01	0.877 ± 0.01	0.151 ± 0.01	0.094 ± 0.00	0.005 ± 0.00	0.055 ± 0.02	0.149 ± 0.12	4.125
	RAC (AAAI'26)	0.869 ± 0.01	0.092 ± 0.01	0.082 ± 0.05	0.357 ± 0.06	0.142 ± 0.01	0.083 ± 0.01	0.005 ± 0.00	0.082 ± 0.00	0.199 ± 0.13	6.375
	SubTab (NeurIPS'21)	0.745 ± 0.05	0.086 ± 0.03	0.195 ± 0.08	0.885 ± 0.01	0.241 ± 0.05	0.091 ± 0.01	0.006 ± 0.00	0.154 ± 0.02	0.158 ± 0.04	5.875
	ICE-T (AAAI'25)	**0.904** ± 0.02	0.024 ± 0.04	0.001 ± 0.00	0.862 ± 0.02	0.318 ± 0.04	**0.112** ± 0.00	0.000 ± 0.00	0.040 ± 0.00	0.248 ± 0.02	4.750
	T-JEPA (ICLR'25)	0.873 ± 0.03	0.003 ± 0.00	0.003 ± 0.00	0.819 ± 0.00	0.335 ± 0.03	0.098 ± 0.00	0.008 ± 0.00	0.143 ± 0.01	0.150 ± 0.10	7.000
	FUSE (Ours)	0.882 ± 0.01	**0.142** ± 0.01	**0.230** ± 0.03	**0.950** ± 0.01	**0.390** ± 0.02	0.102 ± 0.01	**0.099** ± 0.01	0.109 ± 0.01	**0.324** ± 0.04	**1.750**
ARI	OH-KM	0.693 ± 0.08	0.186 ± 0.00	0.256 ± 0.03	0.706 ± 0.01	0.002 ± 0.00	0.023 ± 0.01	−0.018 ± 0.00	0.136 ± 0.00	0.232 ± 0.00	7.000
	K-Pro. (PAKDD'97)	0.588 ± 0.00	0.092 ± 0.04	0.194 ± 0.08	0.693 ± 0.14	0.002 ± 0.02	0.013 ± 0.01	−0.016 ± 0.00	0.121 ± 0.05	0.221 ± 0.03	8.875
	HARR (ESWA 2025)	0.649 ± 0.06	0.075 ± 0.09	0.023 ± 0.01	0.634 ± 0.19	−0.015 ± 0.02	0.012 ± 0.02	−0.003 ± 0.00	0.113 ± 0.02	0.259 ± 0.15	8.625
	HyReaL (DASFAA'26)	0.769 ± 0.03	0.137 ± 0.00	0.055 ± 0.07	0.582 ± 0.08	0.004 ± 0.01	0.029 ± 0.01	–	–	–	7.500
	SCM (IJCAI'24)	0.844 ± 0.09	0.148 ± 0.03	0.021 ± 0.01	0.842 ± 0.07	0.118 ± 0.02	0.111 ± 0.03	−0.002 ± 0.03	0.126 ± 0.08	0.097 ± 0.02	4.750
	MGBCC (AAAI'25)	0.841 ± 0.07	0.213 ± 0.01	0.075 ± 0.02	0.879 ± 0.00	0.012 ± 0.01	0.109 ± 0.00	−0.007 ± 0.00	0.108 ± 0.05	0.116 ± 0.12	3.875
	RAC (AAAI'26)	0.850 ± 0.01	0.181 ± 0.03	0.144 ± 0.07	0.209 ± 0.06	0.008 ± 0.00	0.054 ± 0.01	−0.010 ± 0.00	0.166 ± 0.00	0.212 ± 0.13	5.625
	SubTab (NeurIPS'21)	0.632 ± 0.12	0.157 ± 0.07	0.291 ± 0.13	0.876 ± 0.01	0.094 ± 0.04	0.057 ± 0.00	−0.014 ± 0.02	0.175 ± 0.00	0.257 ± 0.06	5.875
	ICE-T (AAAI'25)	0.874 ± 0.01	0.034 ± 0.08	−0.002 ± 0.00	0.722 ± 0.01	0.428 ± 0.03	0.066 ± 0.01	0.002 ± 0.00	0.084 ± 0.01	0.317 ± 0.03	5.750
	T-JEPA (ICLR'25)	0.797 ± 0.12	−0.003 ± 0.00	0.007 ± 0.01	0.701 ± 0.00	0.138 ± 0.03	0.064 ± 0.01	0.014 ± 0.03	0.181 ± 0.01	0.200 ± 0.06	7.000
	FUSE (Ours)	**0.884** ± 0.02	**0.245** ± 0.02	**0.309** ± 0.04	**0.958** ± 0.01	**0.235** ± 0.02	**0.133** ± 0.01	**0.212** ± 0.02	**0.202** ± 0.01	**0.368** ± 0.02	**1.125**

Table 1: Overall clustering performance (mean ± std). The best and second-best results are highlighted in bold and underlined, respectively. '–' indicates Out-Of-Memory, and Avg.Rank denotes the average rank across all datasets.

	Variant	ZOO	BC	HF	DG	ANL	DRUG	ADT	NOM
ARI	Full	**0.884** ± 0.02	**0.245** ± 0.02	**0.309** ± 0.04	**0.958** ± 0.01	**0.235** ± 0.02	**0.133** ± 0.01	**0.202** ± 0.01	**0.368** ± 0.02
	w/o TAF	0.729 ± 0.01	0.231 ± 0.03	0.246 ± 0.04	0.778 ± 0.01	0.027 ± 0.02	0.048 ± 0.00	0.027 ± 0.03	0.002 ± 0.00
	w/o CRU	0.836 ± 0.05	0.180 ± 0.01	0.252 ± 0.01	0.752 ± 0.04	0.155 ± 0.03	0.057 ± 0.00	0.186 ± 0.00	0.116 ± 0.02
	w/o Both	0.698 ± 0.00	0.131 ± 0.08	0.250 ± 0.11	0.151 ± 0.06	0.008 ± 0.01	0.026 ± 0.01	0.029 ± 0.04	0.002 ± 0.00

Table 2: ARI performance of FUSE and three ablated variants on eight datasets. Results are reported as mean ± std. The best result for each dataset and metric is highlighted in bold.

4.3 Ablation Studies

To isolate the contributions of TAF and CRU, Table 2 reports a matched 2×2 ablation comparing Full, w/o TAF, w/o CRU, and w/o Both across all datasets. All variants use matched training settings, and results are reported as mean ± population standard deviation of ARI.

Full achieves the highest mean ARI on nine of the ten datasets, with only a near-tied result on CMC. Removing TAF causes pronounced degradation on ZOO, ANL, ADT, and NOM, indicating that type-specific branch reconstruction is important for preserving heterogeneous semantics before fusion. Removing CRU is particularly harmful on DG, BC, and NOM, suggesting that target-exclusive prediction captures useful cross-attribute dependencies. Removing both components leads to the largest performance loss on datasets such as DG, AC, ANL, and DRUG, supporting their complementary roles in type-faithful representation formation and relational unification.

4.4 Convergence Analysis

Training dynamics are evaluated by recording unified-space ACC, NMI, ARI, and training loss at every epoch on datasets DG, ZOO, and NOM. The three datasets represent a mixed-type case with substantial ordinal structure, a nominal-only case, and a large-scale mixed-type case, respectively.

As shown in Figure 3, the training loss consistently follows an overall downward trend. After an initial adjustment period, the three clustering metrics become relatively stable by approximately 20 epochs across the three evaluated datasets.

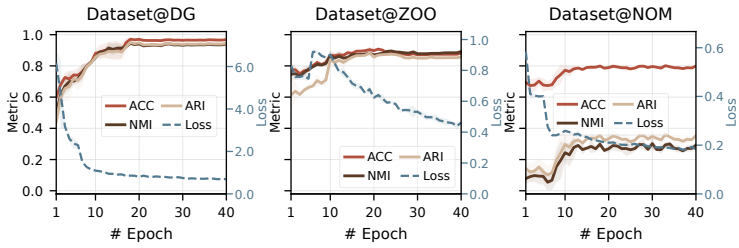

Figure 3: Training loss and ACC, NMI, and ARI over epochs on DG, ZOO, and NOM.

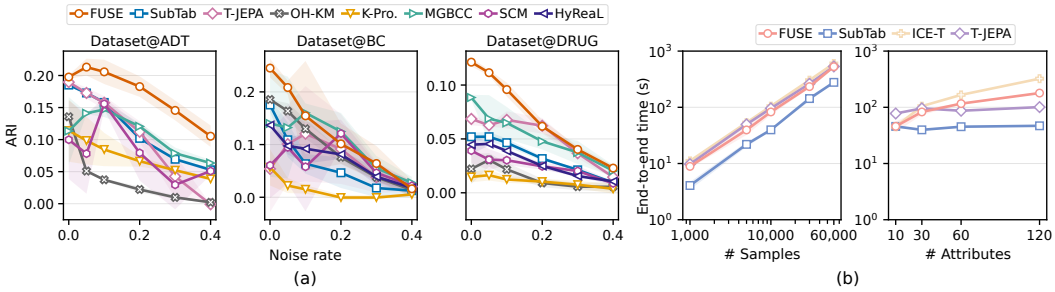

Figure 4: Robustness and runtime scalability of FUSE. (a) ARI under attribute-wise marginal-replacement noise with replacement probability ρ, compared with seven baselines. (b) End-to-end runtime versus the number of samples and attributes.

4.5 Robustness and Scalability

To assess robustness to attribute corruption, each feature cell is independently replaced by a sample from its empirical marginal distribution with probability $\rho \in \{0, 0.05, 0.10, 0.20, 0.30, 0.40\}$. All methods use the same corruption masks and donor observations. As shown in Figure 4(a), FUSE performs best at low-to-moderate noise levels on ADT and DRUG and remains competitive on BC, while SubTab and T-JEPA also retain useful performance. FUSE's stronger robustness is consistent with the complementary type-matched recovery of TAF and cross-attribute prediction of CRU.

To examine runtime scaling with the sample size N and attribute dimensionality p, we use synthetic mixed-type datasets with five balanced clusters and a fixed composition of 50% numerical, 30% nominal, and 20% ordinal attributes. The sample-size sweep varies $N \in \{1\text{k}, 5\text{k}, 10\text{k}, 30\text{k}, 60\text{k}\}$ at $p = 30$, while the attribute-size sweep varies $p \in \{10, 30, 60, 120\}$ at $N = 10\text{k}$. Figure 4(b) reports the end-to-end runtime. FUSE exhibits an empirical runtime exponent of 0.991 with respect to N, indicating near-linear scaling over the evaluated range. Runtime also increases smoothly with p, consistent with the linear dependence on the number of attributes in the TAF and CRU computation.

5 Conclusion

This paper presents FUSE, an attribute-centric self-supervised framework for clustering mixed-type data. The TAF matches perturbation and recovery objectives to the relational semantics of each attribute type, while the CRU uses target-exclusive contextual prediction to provide cross-attribute supervision for the unified representation. Experiments on eleven benchmark datasets and component ablations support the effectiveness of this design. Despite the effectiveness of FUSE, the current framework focuses on static tabular data with numerical, nominal, and ordinal attributes, and target-wise prediction adds computation as the number of attributes grows. Future work will explore more efficient relational learning, self-supervised pretexts and corruption schemes better suited to heterogeneous attributes, and extensions to more complex settings such as dynamic and temporal data.

AI USE STATEMENT

We used generative AI tools to assist with methodological brainstorming, code implementation and revision, and polishing text. The authors manually reviewed the AI-assisted content and evaluated adopted methodological ideas through experiments. We take responsibility for the final manuscript, code, and reported results.

REPRODUCIBILITY STATEMENT

The main paper describes the model architecture, training objectives, and evaluation procedure. Additional implementation details are provided in the supplementary document. We provide source code and experiment configurations in a ZIP archive submitted as supplementary material.

REFERENCES

Amir Ahmad and Lipika Dey. A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set. *Pattern Recognition Letters*, 28 (1):110–118, 2007.

YM Asano, C Rupprecht, and A Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations*, 2020.

Dara Bahri, Heinrich Jiang, Yi Tay, and Donald Metzler. SCARF: self-supervised contrastive learning using random feature corruption. In *Proceedings of the International Conference on Learning Representations*, 2022.

Adrien Bardes, Jean Ponce, and Yann Lecun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR 2022-International Conference on Learning Representations*, 2022.

Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 132–149, 2018.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

Junyang Chen, Yuzhu Ji, Rong Zou, Yiqun Zhang, and Yiu-ming Cheung. Qgrl: Quaternion graph representation learning for heterogeneous feature data clustering. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 297–306, 2024.

Junyang Chen, Yang Lu, Mengke Li, Cuie Yang, Yiqun Zhang, and Yiu-Ming Cheung. HyReaL: clustering attributed graph via hyper-complex space representation learning. In *Proceedings of the International Conference on Database Systems for Advanced Applications*, 2026. in press.

Yiu-ming Cheung and Hong Jia. Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number. *Pattern Recognition*, 46(8):2228–2238, 2013.

Carmelo Del Coso, Diego Fustes, Carlos Dafonte, Francisco J. Nóvoa, José M. Rodríguez-Pedreira, and Bernardino Arcay. Mixing numerical and categorical data in a self-organizing map by means of frequency neurons. *Applied Soft Computing*, 36:246–254, 2015.

Shihao Dong, Yue Liu, Xiaotong Zhou, Yuhui Zheng, Huiying Xu, and Xinzhong Zhu. RAC-DMVC: Reliability-aware contrastive deep multi-view clustering under multi-source noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 20835–20843, 2026.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2019. URL https://archive.ics.uci.edu.

Eibe Frank and Mark Hall. A simple approach to ordinal classification. In *Machine Learning: ECML 2001: 12th European Conference on Machine Learning, Freiburg, Germany, September 5-7, 2001. Proceedings*, volume 2167, pp. 145. Springer, 2003.

J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4): 857–871, 1971.

Zhexue Huang. Clustering large data sets with mixed numeric and categorical values. In *Proceedings of the First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 21–34. World Scientific, 1997.

Zhexue Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3):283–304, 1998.

Lynette Hunt and Murray Jorgensen. Clustering mixed data. *WIREs Data Mining and Knowledge Discovery*, 1(4):352–361, 2011. doi: 10.1002/widm.33.

Dino Ienco, Ruggero G. Pensa, and Rosa Meo. Context-based distance learning for categorical data clustering. In *Proceedings of the International Symposium on Intelligent Data Analysis*, volume 5772, pp. 83–94, 2009.

Anil K. Jain, M. Narasimha Murty, and Patrick J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999. doi: 10.1145/331499.331504.

Si Quang Le and Tu Bao Ho. An association-based dissimilarity measure for categorical data. *Pattern Recognition Letters*, 26(16):2549–2557, 2005.

Kyungeun Lee, Ye Seul Sim, Hyeseung Cho, Moonjung Eo, Suhee Yoon, Sanghyu Yoon, and Woohyung Lim. Binning as a pretext task: Improving self-supervised learning in tabular domains. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 26929–26947, 2024.

Caixuan Luo, Jie Xu, Yazhou Ren, Junbo Ma, and Xiaofeng Zhu. Simple contrastive multi-view clustering with data-level fusion. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 4697–4705, 2024.

Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4920–4928, 2016.

Gowthami Somepalli, Micah Goldblum, Avi Schwarzschild, C. Bayan Bruss, and Tom Goldstein. SAINT: improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*, 2021. URL https://arxiv.org/abs/2106.01342.

Peng Su, Shudong Huang, Weihong Ma, Deng Xiong, and Jiancheng Lv. Multi-view granular-ball contrastive clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20637–20645, 2025. doi: 10.1609/aaai.v39i19.34274.

Hugo Thimonier, José Lucas de Melo Costa, Fabrice Popineau, Arpad Rimmel, and Bich-Liên Doan. T-JEPA: Augmentation-free self-supervised learning for tabular data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.05016.

Tomas Tokar and Scott Sanner. ICE-T: Interactions-aware cross-column contrastive embedding for heterogeneous tabular datasets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20904–20911, 2025. doi: 10.1609/aaai.v39i20.35385.

Talip Ucar, Ehsan Hajiramezanali, and Lindsay Edwards. SubTab: Subsetting features of tabular data for self-supervised representation learning. *Advances in Neural Information Processing Systems*, 34, 2021.

Joaquin Vanschoren, Jan N. van Rijn, Bernd Bischl, and Luis Torgo. OpenML: Networked science in machine learning. *ACM SIGKDD Explorations Newsletter*, 15(2):49–60, 2014. doi: 10.1145/2641190.2641198.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pp. 478–487. PMLR, 2016.

Jinsung Yoon, Yao Zhang, James Jordon, and Mihaela van der Schaar. VIME: extending the success of self- and semi-supervised learning to tabular domain. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pp. 12310–12320. PMLR, 2021.

Yiqun Zhang and Yiu-ming Cheung. Graph-based dissimilarity measurement for cluster analysis of any-type-attributed data. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9): 6530–6544, 2023.

Yiqun Zhang, Yiu-ming Cheung, and Kay Chen Tan. A unified entropy-based distance metric for ordinal-and-nominal-attribute data clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 31(1):39–52, 2020.

Yiqun Zhang, Yiu-ming Cheung, and An Zeng. Het2Hom: representation of heterogeneous attributes into homogeneous concept spaces for categorical-and-numerical-attribute data clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3758–3765, 2022.

Yiqun Zhang, Mingjie Zhao, Yizhou Chen, Yang Lu, and Yiu-ming Cheung. Learning unified distance metric for heterogeneous attribute data clustering. *Expert Systems with Applications*, 273: 126738, 2025a. doi: 10.1016/j.eswa.2025.126738.

Yunfan Zhang, Xiaopeng Luo, Qingsheng Chen, Rong Zou, Yiqun Zhang, and Yiu-ming Cheung. Towards unbiased minimal cluster analysis of categorical-and-numerical attribute data. In *Proceedings of the International Conference on Pattern Recognition*, pp. 254–269, 2024.

Yunfan Zhang, Rong Zou, Yiqun Zhang, Yue Zhang, Yiu-ming Cheung, and Kangshun Li. Adaptive micro partition and hierarchical merging for accurate mixed data clustering. *Complex & Intelligent Systems*, 11(1):1–14, 2025b.

Chengzhang Zhu, Longbing Cao, and Jianping Yin. Unsupervised heterogeneous coupling learning for categorical representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):533–549, 2022.