

# Cubism: Co-Balanced Mixup for Unsupervised Volcano-Seismic Knowledge Transfer

Mahsa Keramati (✉)<sup>1</sup>, Mohammad A. Tayebi<sup>1</sup>, Zahra Zohrevand<sup>1</sup>, Uwe Glässer<sup>1</sup>, Juan Anzieta<sup>2</sup>, and Glyn Williams-Jones<sup>2</sup>

<sup>1</sup> School of Computing Science

<sup>2</sup> Department of Earth Sciences

Simon Fraser University

{mahsa\_keramati,tayebi,zahra\_zohrevand,glaesser,janzieta,glynwj}@sfu.ca

**Abstract.** Volcanic eruptions are severe global threats. Forecasting these unrests via monitoring precursory earthquakes is vital for managing the consequent economic and social risks. Due to various contextual factors, volcano-seismic patterns are not spatiotemporal invariant. Training a robust model for any novel volcano-seismic situation relies on a costly, time-consuming and subjective process of manually labeling data; using a model trained on data from another volcano-seismic setting is typically not a viable option. Unsupervised domain adaptation (UDA) techniques address this issue by transferring knowledge extracted from a labeled domain to an unlabeled one. A challenging problem is the inherent imbalance in volcano-seismic data that degrades the efficiency of an adopted UDA technique. Here, we propose a co-balanced UDA approach, called **Cubism**, to bypass the manual annotation process for any newly monitored volcano by utilizing the patterns recognized in a different volcano-seismic dataset with labels. Employing an invertible latent space, **Cubism** alternates between a co-balanced generation of semantically meaningful inter-volcano samples and UDA. Inter-volcano samples are generated via the mixup data augmentation technique. Due to the sensitivity of mixup to data imbalance, **Cubism** introduces a novel co-balanced ratio that regulates the generation of inter-volcano samples considering the conditional distributions of both volcanoes. To the best of our knowledge, **Cubism** is the first UDA-based approach that transfers volcano-seismic knowledge without any supervision of an unseen volcano-seismic situation. Our extensive experiments show that **Cubism** significantly outperforms baseline methods and effectively provides a robust cross-volcano classifier.

**Keywords:** volcano-seismic event classification · unsupervised domain adaptation · imbalanced data · mixup · flow-based generative models.

## 1 Introduction

Monitoring volcanic unrest is a topic of significant interest, with volcanic hazards threatening the lives of more than 800 million people who live in the vicinity of active volcanoes [22]. Volcanic activity originates from physical processes related to fluid and energy transportation and ranges from gas or non-explosive

lava emissions to extremely violent explosive bursts that may last many hours. These activities generally lead to breaks or cracks in rocks that surge seismicity beneath a volcano before an eruption [22]. Seismological observations are of vital importance for volcanic monitoring as they provide real-time internal data about volcano-seismic events with a high temporal resolution [22].

Rapidly increasing volumes of recorded seismic data call for a paradigm shift in volcano monitoring and forecasting the associated risks. While conventional manual solutions are not practical anymore due to their time-consuming and resource-intensive nature, advanced machine learning techniques and automated predictive analytics can assist risk mitigation and management by annotating seismic events in supervised [4, 19] and unsupervised [3, 7] manners. Unsupervised approaches are prone to low performance and applications of supervised methods are typically limited [2]. The reason is that supervised techniques suffer from low generalization power due to the challenges of unifying the data characteristics from different sources. The unification process is not trivial because of the contextual factors such as soil characteristics and source geometry in addition to potential noise introduced during signal recording [2]. Several methods [2, 3] are proposed to generate a unified feature space for seismic events pursuing a general purpose solution; however, their unification process is highly subjective to signal standardization since these methods alleviate noise impact by relying on a manual selection of intrinsic mode function (IMF) components. This process disregards important aspects of domain shift. Note that the performance of these methods still depends on availability of several large labeled volcano-seismic datasets.

In this paper, we propose **Cubism**, an algorithmic method for robust contextual unification and effective knowledge transfer in the volcano-seismic domain. **Cubism** is an unsupervised domain adaptation (UDA) approach that effectively alleviates the negative impact of inherent class imbalance in volcano-seismic domains by introducing a novel co-balanced inter-volcano modeling. UDA imposes domain-invariance by mitigating the shift between the data distributions of a labeled and an unlabeled domain. Despite being a well-studied area of research, mitigating significant domain gaps is still challenging [20]. Recently, UDA approaches [20, 26] were proposed to alleviate this gap by continuously modeling inter-domain latent space using an emerging vicinal risk minimization technique known as mixup training [28]. These approaches produce inter-domain samples through convex combinations of data and labels/pseudo-labels on input or latent manifold. Yet, these methods don't guarantee semantically meaningful inter-domain modeling [27] and are also subject to bias in case of skewed class distributions. **Cubism** addresses these two issues using a novel co-balanced mixup in the latent space of the flow-based Gaussian mixture model (FlowGMM) [10]. As a result, our proposed solution is capable of developing a robust UDA for volcano-seismic knowledge transfer.

**Cubism** employs FlowGMM because it encourages semantically meaningful inter-domain modeling through a sequence of invertible transformations as a characteristic of flow-based generative models [27]. In addition, FlowGmm as-

sists with conditional mixup by providing a disentangled latent representation. The conditional generative model learned by FlowGMM accommodates inter-domain mixup concerning class imbalance in the labeled domain. Re-balancing that only considers the labeled domain disregards the conditional distribution of the unlabeled domain [25]. To address this issue, **Cubism** proposes co-balanced mixup to address bias and reverse-bias utilizing pseudo-labels of the unlabeled domain; it imposes a robust discriminative cross-volcano feature space with an interplay between a co-balanced mixup and an adversarial UDA.

Co-balanced mixup, first generates samples from a disentangled representation learned by FlowGMM. Then, it models inter-domain space by linear interpolation of these generated samples and samples from the unlabeled domain, considering the conditional skewness of both domains. Finally, for the adversarial adaptation, the training data and inter-domain samples are fed to a soft domain discriminator and a soft classifier so that the minimax game between the discriminator and flow model mitigates the domain gap smoothly [26].

To the best of our knowledge, **Cubism** is the first work that proposes unsupervised volcano-seismic knowledge transfer by employing unsupervised cross-domain classification. To evaluate **Cubism**, we use two real-world volcano-seismic datasets (see Sect. 3.2) from the Llaima and Deception Island volcanoes located in Chile and Antarctica. We define two unsupervised knowledge transfer tasks from each dataset to the other one and assess the cross-volcano classification performance of **Cubism** and several baselines. Our experimental results confirm the effectiveness of **Cubism** through a comparative study where **Cubism** significantly outperforms the strongest baseline by 9.4% in terms of F1-score.

## 2 Related Work

**Volcano-seismic data analysis.** Many works in the literature addressed the classification task of volcano-seismic events utilizing anomaly detection or other machine learning approaches [23, 24]. Despite their promising performance for a specific volcano-seismic situation, they fail to generalize well to a different temporal or spatial volcano-seismic domain [3, 19]. Limited works tackled the problem of designing a spatio-temporal invariant volcano-seismic recognition system. For example, [1] proposes a Bayesian-based approach to learn the mixture of events from two different volcanoes, or [3] trains a model on the standardized data from several volcanoes. The mentioned efforts not only expect large labeled data from multiple volcanoes but also do not aggregate the collected contexts into a unified contextual representation. These problems motivate others to apply unsupervised learning on unlabeled volcano-seismic datasets at the expense of accuracy [3, 7]. In this work, we aim for a more challenging yet rewarding task, unsupervised knowledge transfer from a volcano with annotated data to a different volcano with unlabeled records. One of the only related studies [19] employs active learning to limit the number of required labeled data from a new volcanic setting for training. However, this work does not mitigate the volcano-seismic domain shift; in addition, it requires labeled data from the studied volcanoes and is

subject to class imbalance. Several existing approaches mitigate the scarcity and imbalance in volcano-seismic datasets by applying traditional data augmentation techniques [4] or generative models [9]. However, these approaches are subject to underrepresented samples from low-density regions [15], and their effectiveness are highly dependent on large labeled datasets. Our proposed solution, using an imbalance-aware UDA technique, is the first work to effectively address the above-mentioned issues, to the best of our knowledge.

**Unsupervised domain adaptation (UDA)** is a subcategory of transductive transfer learning that generalizes a model from a labeled source to an unlabelled target under dataset shift. UDA is extensively-studied especially in image processing and computer vision, and existing methods can be categorized into distance-based [12, 20] or adversarial approaches [8, 13, 17]. Distance-based methods train two separate classifiers with some shared layers for each domain. Domain-adversarial neural network (DANN) [8] proposes adversarial UDA as a group of methods that impose domain confusion by a minimax game between a feature extractor and a domain discriminator. Several UDA methods [20, 26] effectively mitigate significant domain gaps by modeling locally Lipschitz in inter-domain space using an efficient regularization technique called mixup [28]. DM-ADA [26] is a mixup-based UDA method that learns the latent representation of both the two domains using variational auto-encoder (VAE) and jointly mixup input and latent space for robust cross-domain classification. Despite promising performances, these state-of-the-art methods are prone to data imbalance issue. Limited UDA approaches address adaptation under imbalanced settings [11]; however, their focus is on label shift that is not aligned with the imbalance problem in volcano-seismic datasets. In this work, we propose **Cubism** as an unsupervised knowledge transfer solution for the volcano-seismic domain considering the data imbalance issue using mixup-based adversarial UDA.

### 3 Basic Concepts and Problem Definition

This section elaborates on the volcano-seismic domain discrepancy problem and its empirical justification based on the characteristics of the datasets employed in our study. Subsequently, we formally define the imbalanced cross-volcano classification problem for volcano-seismic knowledge transfer.

#### 3.1 Volcano-seismic Domain Discrepancy

Each Volcanic hazard has its specific seismic signature. Analyzing the categorical frequency of volcano-seismic activities is a principal step in forecasting volcanic eruptions [22]. Technical quantification of volcanic earthquakes, known as seismic catalog, strengthens volcanic hazards monitoring. Manual detection and labeling of volcanic events by domain experts is not a feasible solution in the era of big data when dealing with massive and rapidly growing volumes of seismic data.

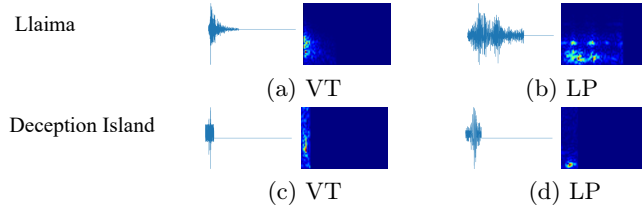


Fig. 1: Volcano-seismic events and their spectrograms

Existing unsupervised methods for seismic event annotation are subject to low performance, and efficient supervised approaches rely on large and high-quality manually labeled data. Still, the manual data annotation process is prone to human subjectivity and lacking unified contextual factors [2, 19, 22]. In addition, characteristics of observed seismicities depend highly on the geophysical properties of volcanoes and placement of sensors [18]. Despite providing high-quality catalogs for signals from a specific station, supervised volcano-seismic recognition (VSR) systems often fail to robustly generalize to a new situation regarding the volcanic state, quality of sensors, environmental noise, etc. [22].

To address the above issues, we propose here to leverage the robustness of the supervised models along with the self-dependency of unsupervised methods. This way, we deliver a model that provides promising catalogs without labeled data for an unseen setting. Utilizing semi-supervised approaches does not effectively fulfill our purpose since these approaches assume the same distribution for labeled and unlabeled data. Therefore, we propose to exploit unsupervised domain adaptation techniques to generalize seismic knowledge from a cataloged volcano-seismic setting to a non-annotated set of events from a different one. Noteworthy, VSR systems suffer from class imbalance issue considering the nature of volcanic activities. Even moderate data imbalance degrades the performance of UDA techniques more than intra-domain learning. Lacking conditional knowledge in unlabeled datasets, the learned cross-domain representation can be biased to the majority classes. None of the explicit data augmentation solutions (undersampling, oversampling, domain-specific approaches, generative models, etc.) efficiently compensate for underrepresented data distributions [28]. Furthermore, although there is an extensive body of research on UDA, just a few recent UDA methods did address the data imbalance. Thus, we propose a novel flow-based method, *Cubism*, that implicitly regularizes adversarial UDA to address imbalanced cross-volcano classification through vicinal risk minimization.

### 3.2 Imbalanced Cross-Volcano Classification

**Exploratory observations.** In this work, we use datasets from two well-studied active volcanoes located in Chile and Antarctica: Llaima [4] and Deception Island [3]. These datasets comprise labeled records of the two most recurrent volcano-seismic events, long period (LP) and volcano-tectonic (VT). Due to their

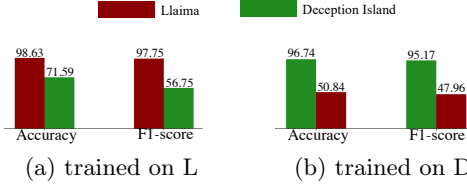


Fig. 2: Cross-volcano generalization power(L: Llama, D: Deception Island)

geophysical origins, signals falling into the same categories share certain characteristics. VT events show an impulsive start and exponential decay, whereas LP signals are non-impulsive and decay slowly. LP events typically are more frequent than VT ones. Despite the common characteristics of all events in the same category, seismicity patterns belonging to different volcanoes do typically not match. Figure 1 illustrates that signals from the Deception Island volcano are nonidentical to signals from the Llama volcano.

To empirically verify volcano-seismic domain discrepancy, we first train a classifier on each dataset and then test it on the other one. As Figure 2 shows, a model trained on one volcano-seismic dataset can successfully be generalized to a test set extracted from the same dataset, while applying the model to another volcano-seismic situation results in poor performance in terms of accuracy and F1-score. This means that leveraging discriminative knowledge from the labeled dataset for annotating the unlabeled one is not feasible without unifying contextual factors.

**Problem definition.** UDA techniques alleviate the cross-domain discrepancy between two different but related domains. We refer to volcano-seismic scenarios with labeled samples as the source domain  $D_s$  and the ones with unlabelled samples as the target domain  $D_t$ , assuming a class imbalance in both  $D_s$  and  $D_t$ . Suppose a given labelled samples  $S = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$  from  $D_s$  along with unlabeled samples  $T = \{(x_i^t)\}_{i=1}^{N^t}$  from  $D_t$ . Our intended task is binary classification:  $D_s$  and  $D_t$  share the same class set ( $C_s = C_t = \{0, 1\}$ ) with label frequencies  $W^s = \{w_0^s, w_1^s\}$  and  $W^t = \{w_0^t, w_1^t\}$ .  $W^t$  is unknown during training and  $w_0^i \neq w_1^i$ , for  $i \in \{s, t\}$ . The objective is to learn the Cubism function that mitigates the domain shift, and given a sample  $x^t$  from  $D_t$ , it accurately predicts label  $y_i^t$ :

$$\text{Cubism}(S, T) \rightarrow \{(y_i^t)\}_{i=1}^{N^t}. \quad (1)$$

## 4 Preliminaries

This section gives an overview of two fundamental concepts our proposed method builds on: mixup regularization and Flow Gaussian Mixture Model.

#### 4.1 Mixup

Mixup [28] is a simple yet effective method to regularize the training process by modelling both intra-class and inter-class vicinity relations. Mixup provides synthetic samples via linear interpolations of pairs of samples and their corresponding labels:

$$x^m = \lambda x_i + (1 - \lambda)x_j ; y^m = \lambda y_i + (1 - \lambda)y_j, \quad (2)$$

where  $(x_i, y_i)$  and  $(x_j, y_j)$  are randomly sampled pairs of (instance, one-hot label) and  $\lambda \in [0, 1]$ .

In the UDA problem, we can mixup (sample, label) pairs from the source domain with (sample, pseudo-label) pairs of the target domain and incorporate these intermediate samples into the adaptation process. This approach improves the efficiency of UDA techniques by modelling inter-domain vicinity relations [26].

#### 4.2 Flow Gaussian Mixture Model (FlowGMM)

A flow-based generative model [5] is an unsupervised model that provides exact inference and density evaluation via seeking an invertible transformation from data space  $X$  to the latent space  $Z$ . This exact mapping from data probability distribution  $P_X$  to a tractable latent probability distribution  $P_Z$  is obtained via the change of variable formula. For the sake of computational simplicity, the latent distribution  $P_Z$  is usually a standard Gaussian.

FlowGMM [10] replaces the Standard Gaussian distribution in the latent space of flow-based models with a Gaussian mixture where each component  $\mathcal{N}(\mu_k, \sigma_k)$  corresponds to class  $k$  in data space  $X$ . This model ( $F_\theta$ ) provides an exact joint likelihood  $P_X(x, y)$  via modelling the exact conditional likelihood  $P(x|y)$  using change of variable formula:

$$L_{GMM} = P_X(x|y = k) = \mathcal{N}(P_Z(\mathcal{F}_\theta(x)|\mu_k, \sigma_k)) \cdot |\det(\frac{\partial \mathcal{F}_\theta(x)}{\partial x})|, \quad (3)$$

and then  $p(y|x)$  is inferred through Bayes' rule as follows:

$$P_X(y|x) = \frac{\mathcal{N}(\mathcal{F}_\theta(x)|\mu_y, \sigma_y)}{\sum_{i=1}^C \mathcal{N}(\mathcal{F}_\theta(x)|\mu_i, \sigma_i)}. \quad (4)$$

### 5 Methodology

We propose Cubism as a robust framework for addressing the imbalanced cross-domain binary classification. Cubism addresses bias and reverse bias in domain alignment through an interplay between *co-balanced inter-domain mixup* and *adversarial UDA* with *conditional mapping*. Figure 3 illustrates the Cubism framework. Each training iteration comprises the following steps:

- (a) Source samples are fed into a FlowGMM  $F$  that maps the complex data space to a latent Gaussian mixture model.

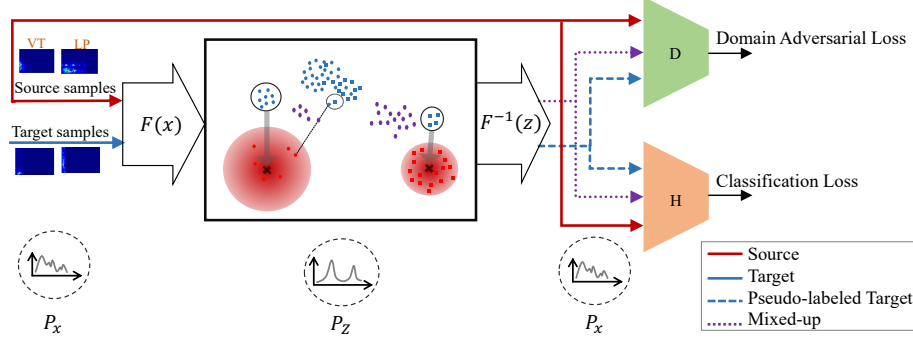


Fig. 3: The architecture of Cubism.

- (b) Target samples are fed into  $F$ , and then corresponding pseudo-labels are assigned to target embeddings based on the learned Gaussians. Note that Cubism estimates the class imbalance ratios in the target domain regarding the conditional information of pseudo-labeled target embeddings and later combines the ratios from both domains producing a set of co-balanced ratios.
- (c) Source-like samples are generated using learned Gaussians applying the co-balanced ratios determined in the previous step.
- (d) Random Gaussian noise is injected to target samples in order to impose local Lipschitz and robustness.
- (e) The generated source samples are linearly interpolated in the latent space along with noisy target embeddings to produce semantically-meaningful inter-domain samples. Invertibility of flow-based models provides a semantic preserving cross-domain data augmentation.
- (f) Source, target and inter-domain samples are fed into an adversarial UDA framework which is slightly modified with a soft domain discriminator and soft source classifier. In addition, for a plausible conditional domain alignment, Cubism maps target samples that are further from the decision boundaries of the source classifier to their corresponding source Gaussians.

Algorithm 1 summarizes Cubism. We elaborate on different aspects of this approach in the following sections.

### 5.1 Co-balanced Inter-Domain Mixup

Co-balanced inter-domain mixup regularizes the inter-domain space in an unbiased manner. Cubism synthesizes inter-domain samples by co-balanced sampling from the source Gaussians and then mixing up the generated samples with the target samples.

**Co-balanced sampling.** We train a FlowGMM model  $F$  on source data since this conditional generative model computes the exact likelihood and maps the

**Algorithm 1:** Co-balanced domain alignment

---

**Input:**  $S = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$ ,  $T = \{(x_i^t)\}_{i=1}^{N^t}$   
**Output:**  $M_{ST}(\{\Psi_F, \Psi_H, \Psi_D\})$ : Final model

```

1  $F = (\Psi_F) \leftarrow preTrain(S)$ ;
2  $W^s \leftarrow classFreq(S)$ ;
3 for  $e \in EPOCH$  do
4   for  $b = \{S_b, T_b\} \in Batches$  do
5      $Z_b^t \leftarrow F(T_b)$ ;
6      $\hat{T}_b \leftarrow softPseudoLabel(Z_b^t)$ ;
7      $\bar{T}_b \leftarrow NoiseInjection(\hat{T}_b)$ ;
8      $\hat{W}_b^t \leftarrow classFreq(\hat{T}_b)$ ;
9     if  $e > 0$  then
10       $\hat{W}^t \leftarrow \xi \hat{W}^t + (1 - \xi) \hat{W}_b^t$ ;
11    else
12       $\hat{W}^t \leftarrow \hat{W}_b^t$ ;
13     $\bar{W} \leftarrow coBalancedFreq(W^s, \hat{W}^t)$ ;
14     $p \leftarrow [1 - \bar{w}_0, 1 - \bar{w}_1]$ ;
15     $Z_b^G \sim G[(\mu_0, \sigma_0), (\mu_1, \sigma_1)]$ ;
16     $\lambda \sim Beta(\alpha, \beta)$ ;
17     $M_b \leftarrow Mixup(Z_b^G, \bar{T}_b, \lambda)$ ;
18     $L_{total} \leftarrow computeLoss(\bar{T}_b, S_b, M_b, \lambda)$ ;
19     $M_{ST}.backpropagate(L_{total})$ ;
20     $M_{ST}.update()$ ;
21 return  $M_{ST}$ 

```

---

source data to a Gaussian mixture model in latent space  $Z$  through optimizing the likelihood explained in Eq. 3:  $Z^s = \{(z^s, Y^s)\}$ ,  $| z^s = F(x^s)$ , where  $Y^s$  is the one-hot label encoding.

Now, to address the class imbalance issue we want to generate samples from the learned Gaussians  $G = [\mathcal{N}(\mu_0, \sigma_0), \mathcal{N}(\mu_1, \sigma_1)]$ . Our empirical observations show that ignoring class imbalance in the target domain misleads the conditional alignment. The reason is that by incorporating imbalanced information from the source, the model gradually will be biased toward the minority class even if the first few training steps follow a normal process. Thus, to dynamically maintain an unbiased mixup, we sample from the set of Gaussians  $G$  considering the class imbalance in both domains.

Class frequency  $W^s = [w_0^s, w_1^s]$  in the source domain is directly estimated via the labels  $Y^s = \{y_i\}_{i=1}^{N^s}$  where  $w_i^s = N_i^s/N^s$ , and  $N_i^s$  is the number of source samples from class  $i$ . However, due to the lack of labels in the target domain, assessing class-imbalance ratios is not as straightforward as it is in the source domain. We estimate these ratios through utilizing obtained pseudo-labels of

the target data. For this, we first map target samples  $T$  to the latent space  $Z$  through network  $F$ , and then, we assign soft pseudo-labels to the embedded target samples representing their classification probability by employing the set of learned Gaussians  $G$ . This process results in a set of softly pseudo-labeled target samples  $\hat{T} = \{(z^t, \hat{Y} = \{\hat{y}_j\}_{j \in \{0,1\}})\}$ , where  $y_j = P_X(y = j|x^t)$  and  $P_X(y = j|x^t)$  is realized through Eq. 4 and  $z^t = F(x^t)$ .

Now, we estimate class ratios in the target domain  $\hat{W}^t = \{\hat{w}_0^t, \hat{w}_1^t\}$  using hard pseudo-labels  $\tilde{Y} = \{\tilde{y}_i\}_{i=1}^{N^t}$  where  $\tilde{y}_i = \underset{j \in \{0,1\}}{\operatorname{argmax}} \hat{Y}_i$ . For a global estimation

of  $\hat{W}^t$ , these weights are adapted during training by accumulating all previous local target ratios:  $\hat{W}^t \leftarrow \xi \hat{W}^t + (1 - \xi) \hat{W}_b^t$ , where  $b$  is the batch number.

Cross-domain imbalance ratio  $\bar{W} = \{\bar{w}_i\}_{i \in \{0,1\}}$  is obtained as follows:

$$\bar{w}_i = \frac{w_i^s + \hat{w}_i^t}{2}. \quad (5)$$

For a co-balanced sampling, we assign sampling probability  $p = \{1 - \bar{w}_0, 1 - \bar{w}_1\}$  to the set of learned Gaussians  $G$ . Then, we use reparameterization trick [14] to generate new source-like samples  $Z^G$  in the latent space  $Z$  regarding the probability set  $p$ :  $Z^G = \{(z^G, y)\}$ ,  $| z^G \stackrel{p}{\sim} \mathcal{N}(\mu_y, \sigma_y)$ ,  $y \in \{0, 1\}$ .

**Manifold mixup.** To bridge the inter-domain gap, we generate intermediary instances via interpolating  $Z^G$  and  $\hat{T}$  and incorporate them to the adaptation process along with source and target data.

Penalizing drastic changes of a classifier prediction affected by input perturbations (a.k.a locally Lipschitz) imposes the cluster assumption [21]. Before mixing up source and target samples, similar to NFM [16] we model locally-Lipschitz by injecting additive and multiplicative noises to the embedded target samples:  $\tilde{z}_i^t = (1 + \sigma_m \zeta_i^m) \cdot z_i^t + \sigma_a \zeta_i^a$ , where  $\zeta_i^m$  and  $\zeta_i^a$  are random variables modeling the desired noise and  $\tilde{T} = \{\tilde{z}_i^t, \hat{Y}_i\}_{i=1}^{N^t}$  is the set of noisy target embeddings.

Now, we randomly mix up generated source samples  $Z^G$  and noisy target samples  $\tilde{T}$ :  $z_i^m = \lambda z_i^G + (1 - \lambda) \tilde{z}_i^t$ ,  $Y_i^m = \lambda Y_i^G + (1 - \lambda) \hat{Y}_i^t$ , where  $\lambda \sim \text{Beta}(\alpha, \beta)$ .

Due to the exact coding and decoding of  $F$ , these linear interpolations in expressive latent space  $Z$  model perceptual mixup on the complex data manifold. As depicted in Figure 3, by using  $M = \{(F^{-1}(z_i^m), Y_i^m)\}_{i=1}^{N^m}$  in addition to  $S$  and  $\tilde{T} = \{(F^{-1}(z_i^t), \hat{Y}_i^t)\}_{i=1}^{N^t}$  we prepare a robust and enriched data as input of adversarial and conditional UDA that can smoothly direct the domain alignment.

## 5.2 Adversarial UDA and Conditional Mapping

The adaptation phase has two main components: a holistic alignment using adversarial UDA and a conditional alignment of target samples with high classification confidence. Holistic adaptation globally aligns the distribution of two domains and conditional alignment enforces discriminative domain transfer.

**Holistic alignment.** Inspired by mixup-based adversarial UDA [26], we modified the DANN [8] architecture to incorporate inter-domain synthetic samples in the adaptation process. DANN is a minimax game between a domain discriminator and a feature extractor, alongside training a classifier for labeled data.

To minimize a cross-entropy loss ( $CE$ ), source samples and their corresponding labels are fed into a two-way classifier  $H$ :

$$L_H = \mathbb{E}_{(x,y) \sim S} [CE(H(G(x)), y)]. \quad (6)$$

Optimizing the classifier  $H$  with respect to the inter-domain samples  $M$  in addition to the source samples encourages locally-Lipschitzness in the inter-domain space. Therefore, inter-domain samples and their soft pseudo-labels are fed to the classifier  $H$  to optimize the following objective:

$$L_H^m = \mathbb{E}_{(x,y) \sim M} [CE(H(G(x)), y)]. \quad (7)$$

For a holistic distribution alignment, a domain label  $l_D$  is assigned to each sample, where  $l_D$  is 1 for  $x \sim S$ , 0 for  $x \sim \tilde{T}$ , and  $\lambda$  for  $x \sim M$ .

Now, source and target samples along with their domain labels are fed to a domain discriminator  $D$  with a classification objective as follows:

$$L_d = \mathbb{E}_{x \sim S} [\log D(G(x))] + \mathbb{E}_{x \sim T} [\log(1 - D(G(x)))]. \quad (8)$$

In addition, we feed inter-domain samples and their domain labels to the domain discriminator to model the inter-domain space:

$$L_d^m = \mathbb{E}_{x \sim M} [l_D \log D(G(x)) + (1 - l_D) \log(1 - D(G(x)))]. \quad (9)$$

Aiming for an adversarial UDA, we maximize domain discriminator losses by optimizing the parameters of the flow model  $F$ , while parameters of  $D$  are trained via minimizing the objective functions of  $D$ . This minimax game imposes global domain confusion; therefore, a domain invariant representation in space  $Z$  is learned through optimizing  $L_H$  and  $L_H^m$  alongside a minimax game for the adversarial loss functions  $L_d$  and  $L_d^m$ :

$$L_{Adv} = \min_{F,H} \max_D \rho L_d + \gamma L_d^m + \eta L_H^m + L_H, \quad (10)$$

where  $\rho$ ,  $\gamma$  and  $\eta$  are hyper-parameters to regulate the interplay of the modules over the course of the adaptation process.

**Conditional alignment.** To impose a plausible conditional alignment, analogous to class-aware UDA [13], we encourage more confident target samples to be aligned with their corresponding class. First, we select a set of easy target samples with a classification confidence higher than a threshold  $\tau$  as follows:

$$T_e = \{(x_e^t, y_e)\} \mid CP(x_e^t, y_e) > \tau, \quad (11)$$

where  $CP(x_e^t, y_e)$  is the probability of sample  $x_e$  belonging to the class  $y$ . Next, we encourage conditional alignment by mapping easy target samples  $T_e$  to their corresponding Gaussians in space  $Z$  by optimizing Eq. 4 as follows:

$$L_c = P_X(x_e^t | y_e = k). \quad (12)$$

Finally, total objective function is:

$$L_{tot} = L_{adv} + \delta L_c = \min_{F,H} \max_D \rho L_d + \gamma L_d^m + \eta L_H^m + L_H + \delta L_c, \quad (13)$$

where  $\delta$  is the regulating factor for conditional alignment. Throughout the training, alternating between the co-balanced mixup and domain alignment for optimizing  $L_{tot}$  effectively align the distribution of the source and target domains. Eventually, this step-by-step process enables the classifier  $H$  to correctly classify samples from the target domain.

## 6 Experiments

We assess here the effectiveness of Cubism on unsupervised volcano-seismic knowledge transfer. We first elaborate on the characteristics of the datasets and the feature extraction process. Then, after briefing implementation setup, we substantiate the efficacy of Cubism via a comprehensive analysis of our experiments.

### 6.1 Data Characteristics

As discussed in Sect. 3.2, we are using event records from Llaima Volcano and Deception Island Volcano as two non-identical volcano-seismic situations. These datasets have pairs of (records segment, event type), where each record segment is a raw stream of seismic signals. The seismic records of Llaima Volcano were captured between 2010 and 2016, while data from the Deception Island Volcano incorporates seismic records belonging to two different periods: 1994-1995 and 2009-2010. Table 1 summarizes the characteristics of these datasets. Note that both datasets suffer from data imbalance issues. To make the signals compatible, we first standardize each record segment with respect to its maximum value and then interpolate the data from Deception Island to match the sampling rate of the data from the Llaima volcano. This process helps to preserve the signals' temporal characteristics. Afterward, by zero-padding the signals, we maintain the same dimension for all the signals. After padding, all the records are set to 6,000 samples (an interval of 60 seconds). Finally, following [4], we utilize short-time fast Fourier transform (FFT) of 512 points to convert the raw datasets to sets of spectrograms. The resulting images are used as input for training Cubism.

### 6.2 Implementation Details

For exact disentangled coding, following [10], we use a RealNVP normalizing flow with two coupling layers, one hidden state and 128 hidden units. Both

Table 1: Characteristics of the studied datasets

Attribute \ Volcano	Llaima	Deception Island
Number of LP events	1310	262
Number of VT events	304	77
Sampling Frequency (Hz)	100	50

source classifier and domain discriminator comprise a backbone, dropout and two fully connected layers with the Relu activation function. A pre-trained Resnet-18 on ImageNet is adopted as the backbone while training the downstream layers from scratch. For training the network, we employ the Adam optimizer with a momentum of 0.9 and a decaying learning rate initiated by 0.01. We set  $\alpha$  and  $\beta$  in Eq. 10 to 8 and 2 respectively,  $\tau$  to 0.9,  $\xi$  to 0.9 and the batch size to 15. Towards an effective interplay of components of **Cubism**, we arrange the values of  $\rho$ ,  $\gamma$ ,  $\eta$  and  $\delta$  (see Eq. 13) to gradually increase with an exponential schedule equal to  $(\frac{2}{1+e^{(-\iota \cdot b)}} - 1)$  [8], where  $\iota = 10$  and  $b$  is increased linearly from 0 to 1.

### 6.3 Empirical Analysis

We define two volcano-seismic knowledge transfer tasks from Llaima Volcano to Deception Island Volcano and from Deception Island Volcano to Llaima Volcano, denoted as  $L \rightarrow D$  and  $D \rightarrow L$ , respectively. To empirically evaluate the efficacy of **Cubism**, we compare the performance of **Cubism** to several baselines on these two tasks. Following is the list of comparison partners in our experiments:

- **Source-Only**: a pre-trained Resnet-18 plus two layers classifier for the tasks  $L \rightarrow D$  and  $D \rightarrow L$  on data from Llaima volcano and Deception Island Volcano, respectively.
- **IMF-STD**: an approach analogous to [2] that standardize volcano-seismic records by using a set of six intrinsic mode function components.
- **DANN** [8]: adversarially aligns distributions disregarding class imbalance.
- **DM-ADA** [26]: a mixup-based adversarial UDA that produces inter-domain samples in input and latent spaces of a VAE without tackling imbalanced data issue.
- **DM-ADA-Flow**: a version of DM-ADA that uses flow-based generative model instead of VAE.
- **BMix-DA**: a version of **Cubism** without addressing the data imbalance issue in the target domain.

**Cross domain classification performance.** Table 2 presents the performance of **Cubism** and the baselines in terms of classification accuracy, precision, recall and F1-score. The results are reported as the average performances over five runs. **Cubism** significantly outperforms all the baselines since it robustly mitigates the domain gap via co-balanced inter-domain modeling. **Cubism** delivers a

Table 2: Methods performance (%) for the transfer tasks of  $L \rightarrow D$  and  $D \rightarrow L$ 

Method	Accuracy		Precision		Recall		F1-score	
	$L \rightarrow D$	$D \rightarrow L$	$L \rightarrow D$	$D \rightarrow L$	$L \rightarrow D$	$D \rightarrow L$	$L \rightarrow D$	$D \rightarrow L$
Source-Only	71.59	50.84	57.46	55.70	56.43	59.10	56.75	47.95
IMF-STD	81.06	71.29	84.46	69.27	59.35	81.17	60.45	67.36
DANN [8]	89.94	91.69	93.25	87.61	78.37	84.27	82.99	85.80
DM-ADA [26]	87.86	88.03	92.01	90.31	73.83	69.39	78.51	74.25
DM-ADA-Flow	88.46	90.51	91.36	92.35	75.59	75.97	80.11	81.04
BMix-DA	93.79	95.23	94.92	96.00	87.28	88.22	90.39	91.50
Cubism	94.98	96.15	95.70	95.75	89.87	91.44	92.38	93.41

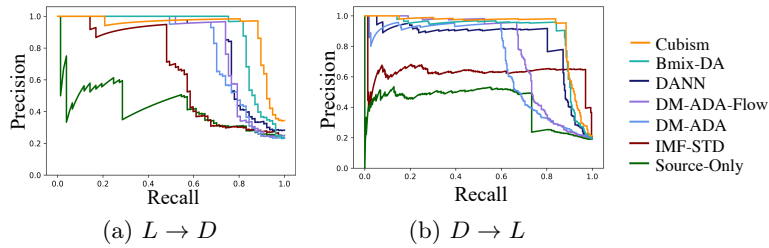


Fig. 4: Precision-Recall curve

more performant model than BMix-DA, confirming the significant contribution of co-balanced mixup in addressing forward bias and reverse bias in the course of training. Besides, BMix-DA outperforms DANN and DM-ADA substantiating the significance of addressing the data imbalance issue using UDA. DM-ADA-Flow replaces the VAE in DM-ADA with a flow-based generative model to assess the effectiveness of an invertible generative model utilization. As shown in Table 2, DM-ADA-flow outperforms DM-ADA due to delivering semantic preserving mixup. Although using Mixup substantially improves the performance of UDA approaches, in the case of data imbalance, these solutions are biased toward the majority class. Thus, as shown in Table 2, DM-ADA has a lower performance compared to DANN. IMF-std is outperformed by all the UDA-based methods confirming the crucial role of UDA techniques in aligning the data distribution of the two volcanoes. Finally, the poor performance of source-only emphasizes the necessity of addressing the inter-volcano gap.

Furthermore, Figure 4 compares the predictive power of the classifier learned by Cubism with the baselines using the Precision-Recall curve. Cubism considerably outperforms all the baselines confirming the performance analysis above.

**Feature visualization.** t-SNE [6] is a widely-used approach to reduce high dimensional data to 2D. For a visualized comparison, we mapped the deep features learned by Source-Only, DANN, DM-ADA and Cubism to 2D space for the transfer task  $L \rightarrow D$  employing t-SNE. Figure 5 demonstrates the t-SNE

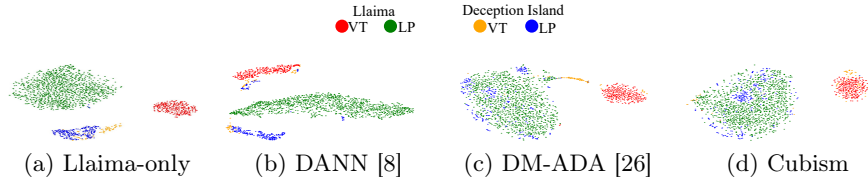


Fig. 5: t-SNE visualization of network activations generated by DANN, DM-ADA and Cubism for the transfer task  $L \rightarrow D$

projections. There is a significant inter-volcano distribution gap for Source-Only as depicted in Figure 5-a. In contrary to DANN and DM-ADA, Cubism is not subject to class imbalance, as one can see in Figures 5b-d. In other words, Cubism imposes the best conditional alignment between the two volcanic domains compared to all baselines.

## 7 Conclusions

Cubism is a novel framework for unsupervised cross-volcano classification that robustly models inter-volcano manifold in an invertible latent space. Cubism goes one step beyond the limited assumption of conditional balance in unsupervised domain adaptation methods by dynamic co-alleviation of bias and inevitable reverse bias. Cubism proposes co-balanced inter-volcano modeling and delivers well-rounded mitigation of the inter-volcano gap. This approach opens a new perspective to significantly less resource-intensive volcano-seismic knowledge transfer with a promising performance. We evaluate Cubism in an extensive comparative study on the knowledge transfer task for two well-studied volcanoes, showing that it outperforms all the baselines by a large margin, thus establishing the efficacy of this new approach. Cubism is a game changer in forecasting volcanic hazards by substantiating a fundamental step toward low-cost mining of volcano-seismic data. Our future work aims at extending Cubism to the more complicated task of unsupervised discriminative knowledge transfer given volcano-seismic stream data where there is no prior knowledge about the similarity of class sets in volcano-seismic situations.

**Acknowledgements** This research has been funded by the Canadian Mountain Network as part of the Networks of Centres of Excellence program and the Natural Sciences and Engineering Research Council of Canada.

## References

1. Bueno, A., et al.: Volcano-seismic transfer learning and uncertainty quantification with bayesian neural networks. *Trans. on Geoscience and Remote Sensing* (2019)
2. Cortés, G., et al.: Practical volcano-independent recognition of seismic events: Vulcan. ears project. *Frontiers in Earth Science* (2021)

3. Cortés, G., et al.: Standardization of noisy volcanoseismic waveforms as a key step toward station-independent, robust automatic recognition. *SLR* (2019)
4. Curilem, M., et al.: Using cnn to classify spectrograms of seismic events from llaima volcano (chile). In: *IJCNN* (2018)
5. Dinh, L., et al.: Nice: Non-linear independent components estimation. *arXiv* (2014)
6. Donahue, J., et al.: Decaf: A deep convolutional activation feature for generic visual recognition. In: *ICML* (2014)
7. Duque, A., et al.: Exploring the unsupervised classification of seismic events of cotopaxi volcano. *Journal of Volcanology and Geothermal Research* (2020)
8. Ganin, Y., et al.: Unsupervised domain adaptation by backpropagation. In: *ICML* (2015)
9. Grijalva, F., et al.: E seismic-gan: A generative model for seismic events from cotopaxi volcano. *J. of Selected Topics in Applied Earth Observations and Remote Sensing* (2021)
10. Izmailov, P., et al.: Semi-supervised learning with normalizing flows. In: *ICML* (2020)
11. Jiang, X., et al.: Implicit class-conditioned domain alignment for unsupervised domain adaptation. In: *ICML* (2020)
12. Kang, G., et al.: Contrastive adaptation network for unsupervised domain adaptation. In: *CVPR*. pp. 4893–4902 (2019)
13. Keramati, M., et al.: Norma: A hybrid feature alignment for class-aware unsupervised domain adaptation. In: *CIKM* (2021)
14. Kingma, D.P., et al.: Auto-encoding variational bayes. In: *ICLR* (2014)
15. Lee, J., et al.: Self-diagnosing gan: Diagnosing underrepresented samples in generative adversarial networks. *Neurips* (2021)
16. Lim, S.H., et al.: Noisy feature mixup. *arXiv* (2021)
17. Madadi, Y., et al.: Deep visual unsupervised domain adaptation for classification tasks: A survey. *Iet Image Processing* (2020)
18. Malfante, M., et al.: Machine learning for volcano-seismic signals: Challenges and perspectives. *Signal Processing Magazine* (2018)
19. Manley, G., et al.: A deep active learning approach to the automatic classification of volcano-seismic events. *Frontiers in Earth Science* (2022)
20. Na, J., et al.: Fixbi: Bridging domain spaces for unsupervised domain adaptation. In: *CVPR* (2021)
21. Shu, R., et al.: A dirt-t approach to unsupervised domain adaptation. In: *ICLR* (2018)
22. Thelen, W.A., et al.: Trends in volcano seismology: 2010 to 2020 and beyond. *Bulletin of Volcanology* (2022)
23. Titos, M., et al.: A deep neural networks approach to automatic recognition systems for volcano-seismic events. *J. of Selected Topics in Applied Earth Observations and Remote Sensing* (2018)
24. Venegas, P., et al.: Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from cotopaxi volcano. *J. of Selected Topics in Applied Earth Observations and Remote Sensing* (2019)
25. Wei, C., et al.: Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In: *CVPR* (2021)
26. Xu, M., et al.: Adversarial domain adaptation with domain mixup. In: *AAAI* (2020)
27. Yüksel, O.K., et al.: Semantic perturbations with normalizing flows for improved generalization. In: *ICCV* (2021)
28. Zhang, H., et al.: mixup: Beyond empirical risk minimization. In: *ICLR* (2018)