Code

Paper

DiffAug: Enhance Unsupervised Contrastive Learning with Domain-Knowledge-Free Diffusion-based Data Augmentation

Zelin Zang**, Hao Luo, Kai Wang, Panpan Zhang, Fan Wang, **Stan Z. Li, Yang You*
AI Lab, Research Center for Industries of the Future, Westlake University

Introduction

Data augmentation methods are either hand-designed or model-based. Hand-designed methods, like color changes and random cropping in visuals or mutations in DNA sequences, require human input and are often data-specific, struggling with complex data where small changes significantly impact semantics. Semantics-independent methods like adding noise exist but aren't always effective. Additionally, hand-designed methods need more samples to mitigate risks from subtle semantic changes, challenging in costly domains like biology. Model-based methods using generative models (VAE, GAN, diffusion) improve training in vision tasks and supervised learning but face concerns about diversity, generalization, and reliance on external data.

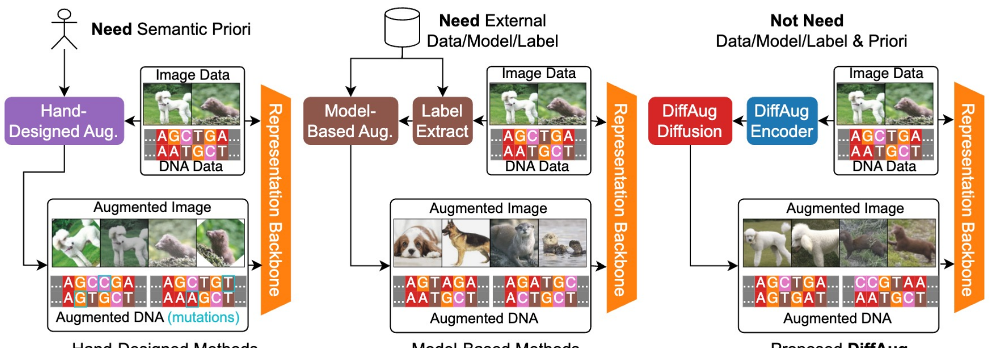
We propose DiffAug, a novel diffusion model-based technique for unsupervised contrastive learning (CL), eliminating the need for training labels. It uses a semantic estimator and a diffusion generator to produce semantically consistent augmented data. DiffAug is effective on DNA, biometric, and visual datasets, outperforming benchmarks in classification and clustering, and operates independently of external data or manual rules.

Methods

Contrastive Learning. Contrastive learning learns visual representation via enforcing the similarity of the positive pairs and enlarging distance of negative pairs. Formally, loss is defined as,

$$-\log \mathcal{Q}\left(\mathbf{z}_i, \mathbf{z}_i^{+}\right) + \log[\mathcal{Q}\left(\mathbf{z}_i, \mathbf{z}_i^{+}\right) + \sum_{\mathbf{z}_i^{-}\in V^{-}} \mathcal{Q}\left(\mathbf{z}_i, \mathbf{z}_i^{-}\right)]$$

Soft Contrastive Learning. To address the performance degradation due to view noise in contrastive learning and to accomplish unsupervised learning on smaller scale datasets (Zang, 2023) designed soft contrastive learning, which soothes sharp positive and negative sample pair labels by evaluating the credibility of the sample pairs. Consider the loss form for multiple positive samples and multiple negative samples.



(a) Hand-Designed Methods (May specific to data domain)

(b) Model-Based Methods (Specific to data domain)

(c) Proposed **DiffAug** (**Not** specific to data domain)

Fig. Comparison of DiffAug with existing augmentation strategy. (a) Hand-designed augmentation is based on human priori that generate new data with different feature but semantically similar semantic. (b) Model-based augmentation methods generate new data with the same labels by training generative models with large amount of data, labels. These methods often require large amounts of data and target specific data domains. (c) DiffAug attempts to reduce the dependence on external data and prior knowledge through iterative training with encoders and diffusion. Expanding the application areas of unsupervised CL.

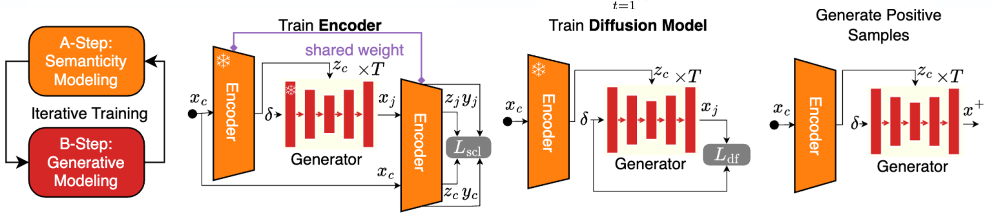
DiffAug Framework. Contrastive learning learns visual representation via enforcing the similarity of the positive pairs and enlarging distance of negative pairs.

Semanticity Modeling (A-Step). In the semanticity modeling step, given a central data x. The $\delta \sim N(0, \mathbf{1})$ is the random initialized data, and z_c is a conditional vector.

$$\mathbf{x}_j = \mathrm{Gen}(\delta, \mathbf{z}_c|\phi^*), \mathbf{y}_c, \mathbf{z}_c = \mathrm{Enc}(\mathbf{x_c}|\theta^*),$$

Generative Modeling (B-Step). In the generative modeling step, the conditional diffusion generator $GEN(\cdot|\phi)$ is trained by the vanilla diffusion loss.

$$\phi = \phi - \eta \sum_{t=1}^{T} \left\{ \left\| \delta - g_\phi\left(\sqrt{\bar{\alpha}_t}\tilde{\mathbf{x}}_c^t + \sqrt{1-\bar{\alpha}_t}, t, \mathbf{z}_c\right) \right\|_2^2 \right\},$$



(a) Training Strategies

(b) Semanticity Modeling (E-Step)

(c) Generative Modeling (M-Step)

(d) Augmentation Generating

Experiments

We conduct experiments on various datasets, including DNA sequences, vision, and bio-feature datasets. We aim to demonstrate that Diffaug can operate effectively and facilitate improvements across diverse domains.

Table 1. **Comparison of Linear probing results on DNA sequence datasets.** The compared methods including SOTA DNA sequence methods (DNA-BERT, NT, Hyena) and contrastive methods with human-designed DNA-augmentation.

Datasets	Genomic Benchmarks (Grešová et al., 2023)			
	MoEnEn	CoIn	HuWo	AVE
CNN	69.0	87.6	93.0	76.7
DNA-BERT	69.4	92.3	96.3	82.5
NT	70.2	90.0	92.3	81.7
Hyena	80.9	89.0	96.4	86.2
SSL+Translocation	83.8	88.2	95.5	80.5
SSL+RC	84.5	88.3	95.8	84.3
SSL+Insertion	80.9	89.8	96.6	85.0
SSL+Mixup	80.9	89.4	96.2	85.4
DiffAug	**86.0**(+1.5)	**94.9**(+2.6)	**96.8**(+0.2)	**89.1**(+2.9)

Table 3. **Comparison of Linear probing results on vision dataset.**

Datasets	CF10	CF100	STL10	TINet
SimCLR	89.6	60.3	89.0	45.2
Mo.V2	86.7	56.1	89.1	47.1
BYOL	92.0	62.7	91.8	46.1
SimSiam	91.6	64.7	89.4	43.0
DINO	91.8	67.4	91.7	44.2
SimC.+Mixup	90.9	62.9	89.6	—
Mo.V2+Mixup	91.5	62.7	90.1	—
SimC.+VAE	89.6	64.2	91.7	46.0
Mo.V2+VAE	89.3	65.9	91.2	43.3
SimC.+GAN	90.0	64.3	89.9	44.6
Mo.V2+GAN	91.1	62.9	91.2	43.6
DiffAug	**93.4**(+1.6)	**69.9**(+2.5)	**92.5**(+0.8)	**49.7**(+2.1)