Federated Self-Supervised Single-cell Clustering of scRNA-seq Data

Shentong Mo
CMU, MBZUAI

Abstract

In recent years, federated self-supervised learning has achieved great progress in the natural language processing and computer vision community. However, little work is exploring self-supervised federated settings on single-cell data, especially on scRNA-seq datasets across various cells. Although one previous work named contrastive-sc on self-supervised single-cell clustering of independently and identically distributed (IID) scRNA-seq data is based on SimCLR-style contrastive learning model, they cannot leverage decentralized unlabeled scRNA-seq data to learn a generic representation with preserving data privacy. To bridge this gap, we introduce a new non-IID scRNA-seq benchmark for federated self-supervised learning to perform single-cell clustering. Furthermore, we propose a novel federated self-supervised learning framework for single-cell clustering, namely FedSC, that can leverage unlabeled data from multiple sequencing platforms to learn scRNA-seq representations while preserving data privacy. We conduct extensive experiments on PBMC & Mouse bladder cells under both IID and non-IID settings. The experimental results demonstrate the effectiveness of our proposed FedSC in federated self-supervised clustering of scRNA-seq data.

1 Introduction

In bioinformatics, single-cell RNA sequencing (scRNA-seq) is a powerful technique for profiling the transcriptomes [1] of individual cells, enabling the discovery of cellular subpopulations [2] and gene expression patterns. The ability to extract crucial biological insights from scRNA-seq data has led to a surge of interest among researchers in analyzing such data.

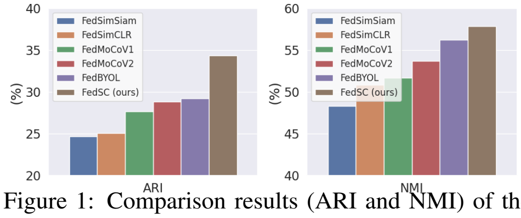

Figure 1: Comparison results (ARI and NMI) of the proposed FedSC with state-of-the-art federated self-supervised baselines (FedSimSiam [3], FedSimCLR [4], FedMoCoV1 [5], FedMoCoV2 [6], and FedBYOL [7]) on mouse bladder cells [8] benchmarks from Microwell-seq platform.

Early scRNA-seq data analysis approaches relied on traditional machine learning techniques, such as Principal Component Analysis (PCA) [9], K-means [10], and Gaussian Mixture Models [11], to conduct cellular subtypes clustering. Due to the challenge of high dimensional and significantly sparse sequences in the clustering analysis, the following methods [12, 13, 14, 15, 16, 17, 18, 19, 20, 21] tried to explore diverse frameworks to address this challenge. Typically, CIDR [12] utilized an implicit imputation stage based on a hierarchical clustering before PCA to alleviate the effect of dropouts. scRNA [14] used a large and well-annotated reference dataset with transferred knowledge by non-negative matrix factorization for small disease-specific data. With the recent advance of deep learning, deep neural networks, such as DCA [17] have been used to boost clustering performance. A clustering layer was adopted in ScDeepCluster [18] on the embedding space learned from DCA to enrich representations. ScziDesk [19] introduced a soft KMeans clustering to aggregate similar

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

cells in the same cluster. However, those methods rely on well-annotated data and take as input the expected number of clusters, which limits their generalization to all circumstances. In contrast, we will solve them in our approach by training a self-supervised learning framework to extract discriminative and compact representations from gene expression inputs of scRNA-seq data.

Recently, inspired by the success of self-supervised learning [5, 4, 6, 22, 23, 24] in image and text, contrastive-sc [25] applied a contrastive loss on anchor and augmented sample outputs from an encoder to extract representations for clustering scRNA-seq data. While the state-of-the-art baseline achieved promising performance, they only focused on self-supervised single-cell clustering of independently and identically distributed (IID) scRNA-seq data via a contrastive learning model based on SimCLR [4], In this case, they cannot leverage decentralized unlabeled scRNA-seq data to learn a generic representation with preserving data privacy. Although some recent federated self-supervised learning frameworks [26, 27] were proposed to learn general representations from images on multiple clients, little work is exploring self-supervised federated settings on single-cell data, especially on scRNA-seq datasets across various cells.

The main challenge is that decentralized gene expression data are growing explosively, and the data collected by multiple parties may not be centralized due to data privacy regulations in real-world scenarios. Utilizing the decentralized unlabeled gene expression data to learn representations with privacy guaranteed is more representative of single-cell clustering. In the meanwhile, previous single-cell clustering approaches [9, 14, 18, 19, 25] are extremely dependent on the assumption that data can be collected and stored in a centralized database, such as gene expression from the sequencing platform. To address the aforementioned challenges, our key idea is to introduce a novel non-IID scRNA-seq dataset for federated self-supervised learning to perform single-cell clustering on the decentralized unlabeled gene expression data, which is different from existing clustering and self-supervised methods. During training, we aim to learn compact representations from multiple clients with decentralized gene expression data across various genes in each cell for discovering potential cellular subtypes.

To this end, we propose a new non-IID scRNA-seq benchmark for federated self-supervised learning to perform single-cell clustering. Furthermore, we present a novel and effective federated self-supervised learning framework for single-cell clustering, namely FedSC, that can leverage unlabeled data from multiple parties to learn general scRNA-seq representations while preserving data privacy. Specifically, our FedSC leverages local self-supervised training on each client to update the server with online communication and aggregation. After aggregation, the server updates the parameters of online encoders in multiple clients. Compared to previous scRNA-seq clustering approaches, our method can extract discriminative scRNA-se representations from decentralized unlabeled gene expression data while preserving data privacy.

We conduct extensive experiments on 10 PBMC & mouse bladder cells under IID and non-IID settings. The experimental results demonstrate the effectiveness of our proposed FedSC in federated self-supervised clustering of scRNA-seq data against the previous centralized scRNA-seq clustering and federated self-supervised learning baselines. Extensive ablation studies also validate the importance of introducing predictor, Exponential Moving Average (EMA), and stop-gradient in federated self-supervised frameworks for learning compact expression representations for federated single-cell clustering of scRNA-seq data. Meanwhile, quantitative comparisons with various numbers of subtypes for each client show the impact of different non-IID levels on federated scRNA-seq clustering.

Our main contributions can be summarized as follows:

- We introduce a new non-IID scRNA-seq benchmark for federated self-supervised learning to perform single-cell clustering.

- We propose a novel federated self-supervised framework for single-cell clustering, namely FedSC, to learn scRNA-seq representations from multiple sequencing platforms while preserving data privacy.

- Extensive experiments on two real scRNA-seq benchmarks comprehensively demonstrate the superiority of our FedSC against the previous centralized scRNA-seq clustering and federated self-supervised learning baselines.

2 Related Work

Federated Learning. Federated learning aims at enabling the training of models without centralizing data, thus preserving user privacy. In federated learning, multiple client devices, such as cell phones and laptops, collaboratively train models by utilizing their local data for training and then uploading their local model updates to the central server for aggregation [28]. In recent years, federated learning has received significant attention [28, 29, 30, 31, 32] and been explored in various application scenarios, such as smartphone keyboard prediction [29], medical image recognition [33], and natural language processing [34]. For instance, some scholars have proposed new aggregation algorithms to improve the accuracy and efficiency of models [31]. Other studies have focused on privacy-preserving techniques, including differential privacy and secure multi-party computation, to protect user data [35]. Furthermore, some studies explored the application of federated learning in specific scenarios, such as mobile intelligence services [36] and edge computing [37]. However, these traditional federated learning approaches rely on data labels, which can not transfer to federated self-supervised single-cell clustering without annotations.

Self-supervised Learning. Self-supervised learning has been addressed in many previous works [4, 38, 22, 5, 6, 3] to learn discriminative representations from internal characteristics of data without any label. Such learnable and transferrable features are beneficial to many downstream tasks, such as image classification [4, 38, 22, 39, 40], object detection [5, 6, 3, 41, 42], semantic textual similarity tasks [23, 24, 43, 44, 45], protein structure prediction [46, 47, 48], and transcription factor binding sites prediction [49, 50, 51]. In the past years, contrastive learning has shown its effectiveness in self-supervised learning, where various instance-wise contrastive learning frameworks [4, 22, 7, 5, 6, 52, 53, 54, 3] and prototype-level contrastive methods [55, 40, 56, 41, 42] were proposed. The general idea of instance-wise contrastive learning is to close the distance of the embedding of different views from the same instance while pushing embeddings of views from different instances away. One common way is to use a large batch size to accumulate positive and negative pairs in the same batch. For instance, Chen *et al.* [4] proposed a simple framework with a learnable nonlinear projection head and a large batch size to improve the quality of the pre-trained representations. Without involving negative instances, BOYL [7] trains the online network from an augmented view of an image to predict the target network representation of the same image under a different augmented view (positive instance). Another broadly-used approach [5] in the self-supervised learning literature is to apply a momentum encoder to update negative instances from a large and consistent dictionary on the fly. In this work, our main focus is to leverage a self-supervised training framework to learn compact gene expression representations from scRNA-seq data for identifying potential cell clusters in federated settings, which is more challenging than the tasks listed above.

Federated Self-supervised Learning. Federated self-supervised learning aims to learn general representations from unlabeled decentralized data while preserving data privacy. In recent years, researchers [57, 58, 26, 27] have tried to explore diverse pipelines to learn discriminative visual representations from decentralized images. For example, FedU [26] utilized a self-supervised framework based on BYOL [7] with a straightforward communication protocol to upload only the weights of online encoders for server aggregation and update them with the aggregated weights. Following up, FedEMA [27] leveraged a divergence-aware decay rate to update the local networks of clients adaptively using the Exponential Moving Average (EMA) of the global network. However, these federated self-supervised learning methods imposed potential privacy leakage risks by directly gathering features and data distribution from clients. Meanwhile, they mainly focused on federated self-supervised learning for image classification. In contrast, we target to address an important but overlooked problem, that is, learn scRNA-seq representations from unlabeled gene expression data from multiple sequencing platforms while preserving data privacy.

Single-cell Clustering. Single-cell clustering of scRNA-seq data is a challenging problem that predicts cellular subtype clusters from gene expression data of diverse cells. Early methods applied classical Principal Component Analysis (PCA) [9], K-means [10], and Gaussian Mixture Models [11] to cluster cell subpopulations from gene expression data directly. Because of the high dimensionality and sparsity of gene expression sequences, the following work [12, 13, 14, 15, 16, 17, 18, 19, 20, 21] explored diverse pipelines to tackle with this issue. For instance, hierarchical clustering with an implicit imputation stage was introduced in CIDR [12] before PCA to address the dropout problem. In order to classify both pure and transitional cells, SOUP [15] utilized the expression similarity matrix to estimate soft membership for cell-type cluster centers. In recent years, deep neural networks, such as DCA [17] have been widely used for extracting expression representations before clustering.

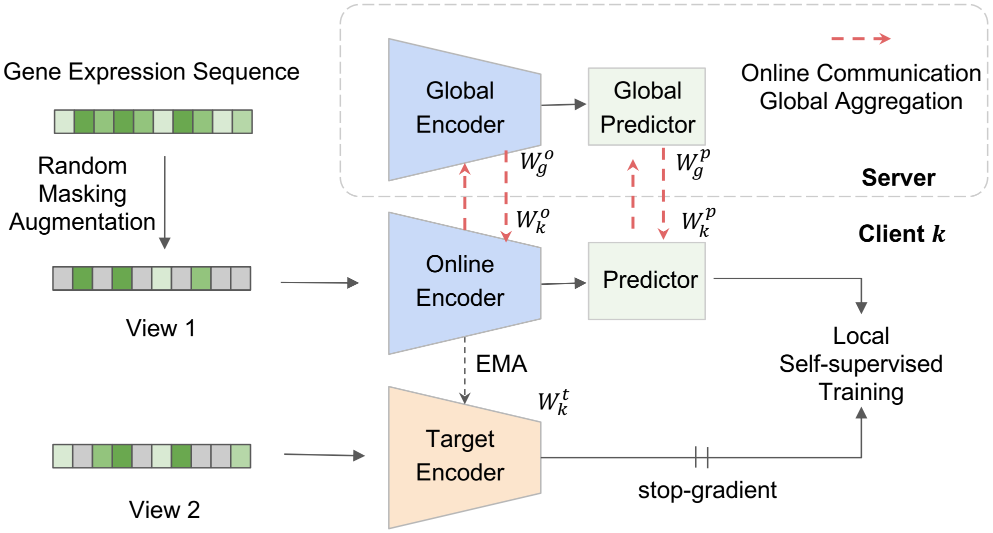

Figure 2: Illustration of the proposed **FedSC** for federated self-supervised single-cell clustering of scRNA-seq data. Each client k takes as input two views of gene expression sequences from random masking augmentation, and conduct local self-supervised training with an online network W_k^o, a predictor W_k^p and an EMA-updated target network W_k^t on unlabeled data $\mathcal{D}_k$. During local training, a stop-gradient operator is applied to the output of the target encoder W_k^t, while the distance between features from the predictor and target encoders is minimized. After local self-supervised training, each client k uploads the trained online models W_k^o and W_k^p to the server and updates it with the global networks W_g^o and W_g^p after aggregation. Finally, the aggregated global encoder W_g^o is used to extract representations of the whole gene expression data to perform single-cell clustering.

Typically, ScDeepCluster [18] used a clustering layer on the embedding space from DCA to enrich embeddings for boosting the clustering performance. ScziDesk [19] proposed a soft self-training KMeans clustering to aggregate similar cells in the same cluster. However, those scRNA-seq clustering baselines mainly rely on well-annotated data, which limits their generalization and violates the fundamental goal of discovering potential cell subtype clusters.

More recently, contrastive-sc [25] introduced the same self-supervised framework as SimCLR [4] to extract embeddings of a short gene expression sequence by an InfoNCE-based contrastive loss. Different from them, we develop a novel multimodal self-supervised framework to learn compact and discriminative representations by reconstructing sequence-level features of masked gene expression matrices for scRNA-seq clustering. However, they cannot leverage decentralized unlabeled scRNA-seq data to learn a generic representation while preserving data privacy. Different from them, we develop a fully novel federated self-supervised learning framework to aggregate unsupervised scRNA-seq representations from the input decentralized gene expression data. To the best of our knowledge, we are the first to introduce new non-IID scRNA-seq benchmarks for federated self-supervised learning to perform single-cell clustering. Our experiments in Section 4.2 also demonstrate the effectiveness of the proposed FedSC in these challenging non-IID settings.

3 Method

Given a set of decentralized gene expression data from scRNA-seq, our target is to learn a global gene expression encoder from multiple clients to extract gene expression embeddings for scRNA-seq clustering. In this work, we propose a federated self-supervised framework for single-cell clustering framework named FedSC for extracting compact and discriminative representations from decentralized single-cell data, which mainly consists of two modules, Local Self-supervised Training for each client in Section 3.2 and Online Communication and Global Aggregation for the server in Section 3.3.

3.1 Preliminaries

In this section, we first describe the problem setup and notations and then revisit contrastive-sc [25], the state-of-the-art baselines for scRNA-seq clustering.

Problem Setup and Notations. Given a set of gene expression data with m genes from n cells, our goal is to learn a discriminative global gene expression feature from m genes in each cell. Note that m, n denote the number of genes and cells, respectively. D denotes the dimension of embeddings. The goal is to learn a generalized representation W from multiple decentralized parties for single-cell clustering. We denote each party as a client k containing unlabeled data $\mathcal{D}_k = \{\mathcal{X}_k\}$. The global objective of federated self-supervised single-cell clustering from multiple parties is $\min_\omega \mathcal{L}(\omega) := \sum_{k=1}^{K} \frac{N_k}{N} \mathcal{L}_k(\omega)$, where $N = \sum_{k=1}^{K} N_k$ denotes the total datasize, and K is the number of clients. For each client k, $\mathcal{L}_k(\omega) := \mathbb{E}_{x_k \in \mathcal{P}_k}[\tilde{\mathcal{L}}_k(\omega; x_k))]$ is the expected object over the gene expression data distribution $\mathcal{P}_k$, where x_k is the gene expression data and $\tilde{\mathcal{L}}_k(\omega; x_k))$ denotes the objective function to train local models from each client.

Revisit contrastive-sc. To solve the single-cell representation learning problem, contrastive-sc [25] first extracted a sequence-level embedding $\mathbf{s} \in \mathbb{R}^{s \times D}$ with a length of s from a gene expression encoder, and then concatenated all m/s sequences as the whole feature for m genes. During training, they utilized a self-supervised framework similar to SimCLR [4] with an InfoNCE-based contrastive loss to close the distance between the anchor and augmented embeddings of one sequence in the same cell, which is denoted as:

$$\mathcal{L}_{\text{contrastive-sc}} = \frac{1}{B} \sum_{i=1}^{B} -\log \frac{\exp\left(\frac{1}{\tau}\texttt{sim}(\mathbf{s}_i, \hat{\mathbf{s}}_i)\right)}{\sum_{j=1}^{B} \exp\left(\frac{1}{\tau}\texttt{sim}(\mathbf{s}_i, \hat{\mathbf{s}}_j)\right)} \tag{1}$$

where $\mathbf{s}_i, \hat{\mathbf{s}}_i \in \mathbb{R}^{1 \times D}$ denote the anchor and augmented embeddings for ith sample in a mini-batch. B is the batch size. $\texttt{sim}(\mathbf{s}_i, \hat{\mathbf{s}}_i) = \mathbf{s}_i^T \hat{\mathbf{s}}_i / (\|\mathbf{s}_i\| \|\hat{\mathbf{s}}_i\|)$ is the cosine similarity, and τ is the temperature parameter. $B^2 - B$ negative sequences are created within a training batch. By optimizing this loss, they successfully extracted discriminative representations of each short sequence in the same cell.

However, this sequence-level contrastive learning framework can not leverage decentralized unlabeled scRNA-seq data to learn a generic representation while preserving data privacy as their framework was trained on centralized gene expression data. To address this issue, we propose a novel federated self-supervised learning framework to learn scRNA-seq representations from decentralized unlabeled data to preserve data privacy, as shown in Figure 2.

3.2 Local Self-supervised Training

In order to explicitly learn unsupervised representation from decentralized gene expression data in different clients, we introduce local self-supervised training on unlabeled data $\mathcal{D}_k$ based on the same global models W_g^o and W_g^p downloaded from the server.

With two views of gene expression sequences from random masking augmentation, we first feed them into each client k with an online network W_k^o, a predictor W_k^p and an EMA-updated target network W_k^t. During local training, a stop-gradient operator is also applied to the output of the target encoder W_k^t for avoiding the local model collapsing problem. Then, the online models are optimized to minimize the distance across output representations from the predictor and target encoders for each client k are minimized as:

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^{B} 2 - 2 * \texttt{sim}(\mathbf{s}_i^p, \mathbf{s}_i^t) \tag{2}$$

where $\mathbf{s}_i^p, \mathbf{s}_i^t$ are gene expression embeddings from the predictor and target encoder for ith sample in a mini-batch. $\texttt{sim}(\mathbf{s}_i^p, \mathbf{s}_i^t) = (\mathbf{s}_i^p)^T \mathbf{s}_i^t / (\|\mathbf{s}_i^p\| \|\mathbf{s}_i^t\|)$ is the cosine similarity. Optimizing the loss for each client k will promote the local online encoder to capture discriminative features of unsupervised gene expression data across different genes for each cell.

3.3 Online Communication and Global Aggregation

With the benefit of local self-supervised learning in each client k, we propose a simple yet effective protocol for online communication and global aggregation between the server and clients. From the

Table 1: Comparison results (%) on 10 PBMC cells and mouse bladder cells benchmarks. ↑ denotes that a large value is better.

Method	PBMC cells			mouse bladder cells		
	ARI (↑)	NMI (↑)	Silhouette (↑)	ARI (↑)	NMI (↑)	Silhouette (↑)
FedSimCLR	51.51	66.30	31.25	25.07	50.85	26.31
FedMoCoV1	51.58	62.62	52.22	27.62	51.68	26.53
FedMoCoV2	55.11	64.99	52.92	28.83	53.70	26.46
FedSimSiam	49.95	60.76	31.99	24.66	48.27	24.89
FedBYOL	59.70	**69.27**	53.12	29.24	56.21	27.56
FedSC (ours)	**62.57**	68.35	**54.70**	**34.35**	**57.84**	**29.58**
PCA (Centralized)	19.00	31.21	40.24	17.28	43.94	21.46
scRNA (Centralized)	47.15	52.91	23.27	23.78	47.68	24.32
contrastive-sc (Centralized)	76.03	76.42	54.96	44.51	69.03	37.72

client to the server, each client k uploads the trained online encoders W_k^o and predictor W_k^p to the server. After aggregation from all clients, the server updates the online encoders W_k^o and predictor W_k^p for each client k with the global model W_g^o and predictor W_g^p.

During training, the online encoder and predictor of each client k at training round r are updated using the EMA of the global network from the server as:

$$\begin{aligned} W_k^{o,r} &= \mu_k W_k^{o,r-1} + (1-\mu_k) W_g^{o,r} \\ W_k^{p,r} &= \mu_k W_k^{p,r-1} + (1-\mu_k) W_g^{p,r} \end{aligned} \tag{3}$$

where $W_k^{o,r}$ and $W_k^{p,r}$ are the parameters of online encoder and predictor of client k at training round r. $W_g^{o,r}$, $W_g^{p,r}$ are the global encoder and predictor of the server. μ_k is the decay rate for each client k that is measured by the divergence between the global $W_g^{o,r}$ and online encoder $W_k^{o,r}$.

Inspired by the insight that retaining local knowledge of non-IID data helps improve performance in a recent federated self-supervised learning framework [27] on decentralized images, we propose to update online networks in each client using different levels of decay rates measured by ℓ_2-norm of the global and online encoders defined as:

$$\mu_k = \min\left(\frac{\alpha}{\|W_g^{o,r} - W_k^{o,r-1}\|}, 1\right) \tag{4}$$

where $\alpha \in [0, 1]$ is a scalar hyper-parameter to control the magnitude of μ at round r. We use $\alpha = 0.7$ in our experiments. Intuitively, we want to retain more local knowledge when the divergence between global and online encoders is large and incorporate more global knowledge when the divergence is small. During inference, we use the global encoder W_g^o to extract gene expression features from input gene expression data for single-cell clustering to generate cellular subtype clusters.

4 Experiments

4.1 Experimental Setup

Datasets. Following scDeepCluster [18], we apply 10 PBMC cells [59] from 10x genomic platform and mouse bladder cell [8] from Microwell-seq for experiments. We use the same split in [25] for training and testing, where the number of cells varies from 870 to 9552, and 4-16 cellular subtype clusters are annotated for evaluation. The 10 PBMC cells [59] dataset with 8 subtype clusters contains 4271 cells and 16653 genes for each cell, while the mouse bladder cells benchmark with 16 subtype clusters includes 2746 cells and 20670 genes per cell.

Evaluation Metrics. Following previous work [25, 18, 19], we apply Adjusted Rand Index (ARI) [60], Normalized Mutual Information (NMI) [61], and Silhouette [62] score for evaluation. ARI score calculates the ratio of sample pairs assigned to the correct cluster labels. NMI score measures the agreement of the ground truth and predicted cluster assignments. A larger value of ARI and NMI is better, which means that the predicted cluster matches the ground-truth cluster. Silhouette score measures the compactness of the generated clusters, and a higher score means that the predicted clusters are denser and better separated.

Table 2: Ablation studies on main components (predictor, EMA, stop-gradient) of federated self-supervised frameworks. Note that FedBYOL is used as the default model.

predictor	EMA	stop-gradient	ARI (↑)	NMI (↑)	Silhouette (↑)
✓	✗	✗	38.16	49.87	37.25
✓	✓	✗	43.57 (**+5.41**)	51.08 (**+1.21**)	42.13 (**+4.88**)
✓	✗	✓	46.21 (**+8.05**)	55.36 (**+5.49**)	45.16 (**+7.91**)
✗	✓	✓	57.63 (**+19.47**)	67.12 (**+17.25**)	51.35 (**+14.10**)
✓	✓	✓	**59.70** (**+21.54**)	**69.27** (**+19.40**)	**53.12** (**+15.87**)

Implementation. Our implementation is based on PyTorch [63] framework. For online and target encoders, the gene expression encoder is composed of 3 linear layers of [200, 40, 60] neurons. The predictor is one linear layer with the output dimension of 128. The model is trained with the SGD optimizer with a learning rate of 0.032, and hyper-parameters with momentum of 0.5 and weight decay of 5e-4. The model is trained for 300 rounds with a batch size of 128, $K = 4$ clients, $E = 5$ local epochs. We use non-IID data with 2 subtype clusters per client for the 10 PBMC benchmark and 4 subtype clusters per client for the Mouse bladder cells dataset. After self-supervised training, we use the K-means algorithm [10] in the scikit-learn package for evaluation to perform clustering on gene expression representations.

4.2 Comparison to prior work

In this work, we propose a novel and effective federated self-supervised training framework for scRNA-seq clustering. To demonstrate the effectiveness of the proposed FedSC, we comprehensively compare it to previous centralized scRNA-seq clustering and federated self-supervised learning baselines: 1) PCA [9]: a traditional machine learning approach with raw gene expression sequence as input to extract principal components; 2) scRNA [14]: a baseline based on non-negative matrix factorization by using transferred knowledge from large and well-annotated data for reference; 3) contrastive-sc [25]: the state-of-the-art self-supervised framework with InfoNCE-based contrastive loss to extract embeddings from sequences only for scRAN-seq clustering; 4) FedSimCLR [4]: a self-supervised baseline with identical encoders and without predictors; 5) FedMoCoV1 [5]: a vanilla self-supervised approach using an online encoder and an EMA-updated target encoder based on memory bank and stop-gradient; 6) FedMoCoV2 [6]: an MoCoV1 improved method by adding a predictor to the online encoder; 7) FedSimSiam [3]: a self-supervised baseline using predictors and stop-gradient; 8) FedBYOL [7]: a strong baseline based on an online encoder with predictors and an EMA-updated target encoder with stop-gradient.

Table 1 reports the quantitative comparison results on 10 PBMC cells and mouse bladder cells datasets. As can be seen, we achieve the best performance in terms of most metrics compared to previous federated self-supervised learning baselines on 10 PBMC cells benchmark. In particular, the proposed FedSC significantly outperforms scRNA [14], the centralized baseline based on non-negative matrix factorization, by 15.42 ARI, 15.44 NMI, and 31.43 Silhouette. Moreover, we achieve superior performance gains of 11.06 ARI, 2.05 NMI, and 23.45 Silhouette compared to FedSimCLR [4], which indicates the importance of predictors and EMA for learning discriminative representations from target encoders in federated self-supervised single-cell clustering. Meanwhile, our FedSC outperforms FedBYOL [7], the current state-of-the-art federated self-supervised approach for scRNA-seq clustering, where we achieve the performance gains of 2.87 ARI and 1.58 Silhouette. These significant improvements demonstrate the superiority of our method in learning compact embeddings from decentralized gene expression data for clustering.

In addition, significant gains in mouse bladder cells benchmark can be observed in Table 1. Compared to FedSimSiam [3], the self-supervised baseline based on predictors and stop-gradient, we achieve the results gains of 9.69 ARI, 9.57 NMI, and 4.69 Silhouette. Furthermore, when evaluated on this challenging benchmark with more cellular subtypes, the proposed approach still outperforms FedBYOL [7] by 5.11 ARI, 1.63 NMI, and 2.02 Silhouette. We also achieve highly better results against FedMoCoV2 [6], the improved MoCoV1 baseline with a predictor added to the online encoder. These results validate the effectiveness of our approach in learning discriminative features from gene expression data in multiple clients for each cellular subtype.

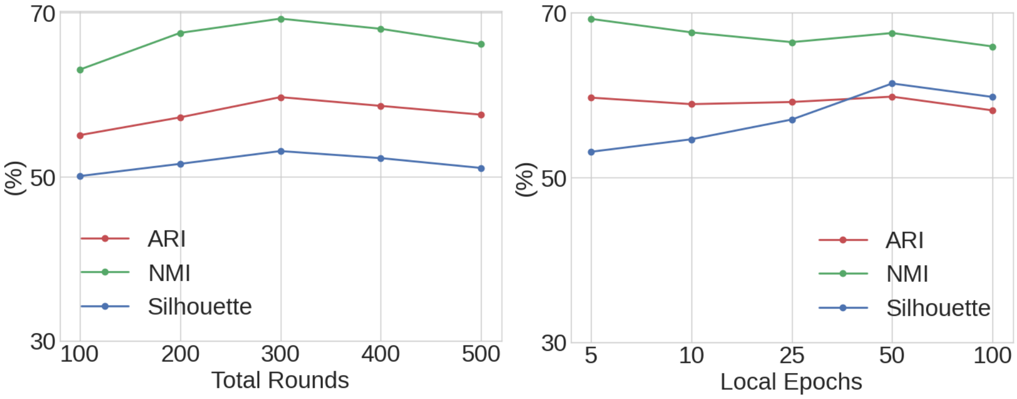

Figure 3: Effect of total rounds and local epochs on the final performance of federated scRNA-seq clustering. FedBYOL achieves the best results at total rounds of 300 and local epochs of 50.

4.3 Experimental analysis

In this section, we performed ablation studies to demonstrate the benefit of introducing three main components (predictor, EMA, stop-gradient) in federated self-supervised frameworks for single-cell clustering. We also conducted extensive experiments to explore the effect of total rounds and local epochs on federated scRNA-seq clustering. Furthermore, we analyze the impact of different non-IID levels on our proposed FedSC method for scRNA-seq clustering.

Ablation on main components (predictor, EMA, stop-gradient). To demonstrate the effectiveness of the introduced predictor, EMA, and stop-gradient for federated self-supervised single-cell clustering, we ablate the necessity of each module and report the quantitative comparison results in Table 2. We can observe that adding EMA to the vanilla baseline with only predictor highly raises the results of scRNA-seq clustering by 5.41 ARI, 1.21 NMI, and 4.88 Silhouette, which validates the benefit of EMA in aggregating local knowledge from online encoders for target encoders to learn discriminative expression representations for discovering accurate cellular subtypes. Meanwhile, introducing the stop-gradient for target encoders in the baseline also increases the clustering performance of all metrics. This indicates that online and target encoders are significantly different during pre-training for scRNA-seq clustering. More importantly, incorporating EMA and stop-gradient for local target encoders together into the baseline significantly raises the performance by 21.54 ARI, 19.40 NMI, and 15.87 Silhouette. Furthermore, we can observe a performance drop without the predictor attached to the local online encoders for feature prediction. These improving results demonstrate the importance of predictor, EMA, and stop-gradient for federated self-supervised learning frameworks in learning discriminative representations from decentralized gene expression data.

Effect of total rounds and local epochs. The number of total rounds and local epochs used in the federated self-supervised learning approach affect the extracted expression representations from gene expression in multiple clients for scRNA-seq clustering. To explore such effects more comprehensively, we ablated the number of total rounds from $\{100, 200, 300, 400, 500\}$ and varied the local epochs from $\{5, 10, 25, 50, 100\}$. The comparison results of federated scRNA-seq clustering are shown in Figure 3. When the number of total rounds is 300 and the number of local epochs is 50, we achieve the best overall clustering performance. With the increase of total rounds from 100 to 300, we can consistently raise results, which shows the importance of training longer time for the server to aggregate compact representation from decentralized gene expression data in each cell. However, increasing the number of total rounds from 300 to 400 and 500 will not continually improve the results of ARI and Silhouette. In particular, a drastic drop can be observed in the NMI score, which means that the generated clusters are not matching with the ground-truth clusters. This might be caused by the high sparsity of non-zero values in the gene expression data. In this case, federated self-supervised single-cell clustering without discriminating this sparsity will deteriorate the quality of representations pre-trained from many zero entries in the input decentralized data.

In terms of local epochs, the performance of the Silhouette score climbs with the increase of the local epochs from 5 to 50. Compared to the Silhouette score, there are no significant changes in ARI and NMI. This interesting trend could be due to the self-property of these metrics. Higher ARI and NMI indicate that the predicted cluster assignment matches the ground-truth cluster assignment,

Table 3: Exploration studies on non-IID levels of federated self-supervised single-cell clustering on mouse bladder cells. Note that our proposed FedSC is used as the default model.

# Subtype Clusters Per Client	ARI (↑)	NMI (↑)	Silhouette (↑)
1	19.57	32.73	14.09
2	30.43	54.01	26.18
4	34.35	57.84	29.58
8	**39.83**	**62.45**	**36.05**

while a larger value of the Silhouette score refers to denser and better-separated clusters. The former metrics can not measure the quality of expression embeddings extracted from the pre-trained gene expression encoder, but the latter is a strict metric for measuring the compactness of learned expression representations for scRNA-seq clustering. Meanwhile, when the number of local epochs is increased to 100, all metrics drop significantly, which might be caused by the limited model capacity of MLP-based encoders for each client to gain more knowledge with more local epochs. For federated self-supervised learning, we are the first to explore such an effect on the server and clients for extracting discriminative features from decentralized gene expression data to conduct clustering.

Impact of non-IID levels. To explore the impact of different non-IID levels on the final performance of federated self-supervised single-cell clustering, we ablated the number of subtype clusters for each client from $\{1, 2, 4, 8\}$. Table 3 reports the comparison results of federated self-supervised scRNA-seq clustering. When the number of non-IID levels is 8, we achieve the best clustering performance in terms of all metrics. With the increase of non-IID levels from 1 to 8, we can consistently raise results, which shows the importance of adding more data in each client for the server to aggregate compact representation from decentralized gene expression data in each cell. This increasing trend is also observed in the recent federated self-supervised learning approaches [27] for image classification. Therefore, how balancing the number of subtype clusters for each client during federated self-supervised single-cell clustering will consistently affect the quality of representations learned from the decentralized gene expression data.

5 Conclusion

In this work, we introduce a novel non-IID scRNA-seq benchmark for federated self-supervised single-cell clustering. Furthermore, we present FedSC, a new federated self-supervised learning framework for single-cell clustering. Our FedSC leverages local self-supervised training on unlabeled data from multiple clients to update the server with online communication and aggregation for learning scRNA-seq representations while preserving data privacy. Extensive experiments on 10 PBMC and Mouse bladder cells demonstrate the effectiveness of our proposed FedSC in federated self-supervised clustering of scRNA-seq data.

Limitation. Although the proposed FedSC achieves superior results on federated single-cell clustering of scRNA-seq data, the performance of non-IID data is a bit far from centralized results. One possible reason is that our model easily overfits the pretext task of federated self-supervised learning during pre-training, and the solution is to incorporate centralized distillation and momentum global encoders together for contrastive expression modeling. The future work could add more decentralized gene expression data for training each client or incorporate continual learning with contrastive expression modeling to increase the compactness of generated clusters.

Broader Impact. The proposed approach successfully learns discriminative representations of decentralized gene expression data across different genes for each cell from manually-collected datasets, which might cause the model to learn internal biases in the data. For instance, the model could fail to discover unseen but crucial cellular subtypes. Therefore, these issues should be addressed for the deployment of real applications in open-world problems.

References

[1] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, et al. mrna-seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5):377–382, 2009. 1

[2] Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. The technology and biology of single-cell rna sequencing. *Molecular cell*, 58(4):610–620, 2015. 1

[3] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 3, 7

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of International Conference on Machine Learning (ICML)*, 2020. 1, 2, 3, 4, 5, 7

[5] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9729–9738, 2020. 1, 2, 3, 7

[6] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 1, 2, 3, 7

[7] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, koray kavukcuoglu, Remi Munos, and Michal Valko. Bootstrap your own latent - a new approach to self-supervised learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1, 3, 7

[8] Xiaoping Han, Renying Wang, Yincong Zhou, Lijiang Fei, Huiyu Sun, Shujing Lai, Assieh Saadatpour, Ziming Zhou, Haide Chen, Fang Ye, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*, 172(5):1091–1107, 2018. 1, 6

[9] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987. 1, 2, 3, 7

[10] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297, 1967. 1, 3, 7

[11] Carl Rasmussen. The infinite gaussian mixture model. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 1999. 1, 3

[12] Peijie Lin, Michael Troup, and Joshua W. K. Ho. Cidr: Ultrafast and accurate clustering through imputation for single cell rna-seq data. *Genome Biology*, 18:59, 2017. 1, 3

[13] Yunpei Xu, Hong-Dong Li, Yi Pan, Feng Luo, and Jianxin Wang. Biorank: A similarity assessment method for single cell clustering. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 157–162. IEEE, 2018. 1, 3

[14] Bettina Mieth, James R F Hockley, Nico Görnitz, Marina M-C Vidovic, Klaus-Robert Müller, Alex Gutteridge, and Daniel Ziemek. Using transfer learning from prior reference knowledge to improve the clustering of single-cell rna-seq data. *Scientific reports*, 9(1):20353, 2019. 1, 2, 3, 7

[15] Lingxue Zhu, Jing Lei, Lambertus Klei, Bernie Devlin, and Kathryn Roeder. Semisoft clustering of single-cell data. *Proceedings of the National Academy of Sciences*, 116(2):466–471, 2019. 1, 3

[16] Xiaoshu Zhu, Lilu Guo, Yunpei Xu, Hong-Dong Li, Xingyu Liao, Fang-Xiang Wu, and Xiaoqing Peng. A global similarity learning for clustering of single-cell rna-seq data. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 261–266. IEEE, 2019. 1, 3

[17] Gökcen Eraslan, Lukas M. Simon, Maria Mircea, Nikola S. Mueller, and Fabian J. Theis. Single cell rna-seq denoising using a deep count autoencoder. *Nature Communication*, 10:390, 2019. 1, 3

[18] Tian Tian, Wan Ji, Song Qi, and Wei Zhi. Clustering single-cell rna-seq data with a model-based deep learning approach. *Nature Machine Intelligence*, 1(4):191–198, 2019. 1, 2, 3, 4, 6

[19] Chen Liang, Wang Weinan, Zhai Yuyao, and Deng Minghua. Deep soft k-means clustering with self-training for single-cell rna sequence data. *NAR Genom Bioinform*, 2020. 1, 2, 3, 4, 6

[20] Ruiqing Zheng, Zhenlan Liang, Xiangmao Meng, Yu Tian, and Min Li. A robust single cell clustering method based on subspace learning and partial imputation. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 140–145. IEEE, 2020. 1, 3

[21] Florian Schmidt and Bobby Ranjan. Robust clustering and interpretation of scrna-seq data using reference component analysis. *bioRxiv*, 2020. 1, 3

[22] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 2, 3

[23] Yan Zhang, Ruidan He, Zuozhu Liu, Kwan Hui Lim, and Lidong Bing. An unsupervised sentence embedding method by mutual information maximization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1601–1610, 2020. 2, 3

[24] Tianyu Gao, Xingcheng Yao, and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6894–6910, 2021. 2, 3

[25] Madalina Ciortan and Matthieu Defrance. Contrastive self-supervised clustering of scrna-seq data. *BMC bioinformatics*, 22(1):1–27, 2021. 2, 4, 5, 6, 7

[26] Weiming Zhuang, Xin Gan, Yonggang Wen, Shuai Zhang, and Shuai Yi. Collaborative unsupervised visual representation learning from decentralized data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4912–4921, 2021. 2, 3

[27] Weiming Zhuang, Yonggang Wen, and Shuai Zhang. Divergence-aware federated self-supervised learning. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2022. 2, 3, 6, 9

[28] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017. 3

[29] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*, 2018. 3

[30] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018. 3

[31] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020. 3

[32] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2020. 3

[33] Georgios A Kaissis, Marcus R Makowski, Daniel Rückert, and Rickmer F Braren. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6):305–311, 2020. 3

[34] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konečnỳ, Stefano Mazzocchi, Brendan McMahan, et al. Towards federated learning at scale: System design. *Proceedings of machine learning and systems*, 1:374–388, 2019. 3

[35] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454–3469, 2020. 3

[36] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3):2031–2063, 2020. 3

[37] Haftay Gebreslasie Abreha, Mohammad Hayajneh, and Mohamed Adel Serhani. Federated learning in edge computing: a systematic survey. *Sensors*, 22(2):450, 2022. 3

[38] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[39] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *Proceedings of International Conference on Machine Learning (ICML)*, 2021. 3

[40] Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. Prototypical contrastive learning of unsupervised representations. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2021. 3

[41] Shentong Mo, Zhun Sun, and Chao Li. Siamese prototypical contrastive learning. In *Proceedings of British Machine Vision Conference (BMVC)*, 2021. 3

[42] Shentong Mo, Zhun Sun, and Chao Li. Rethinking prototypical contrastive learning through alignment, uniformity and correlation. In *Proceedings of British Machine Vision Conference (BMVC)*, 2022. 3

[43] Yuanmeng Yan, Rumei Li, Sirui Wang, Fuzheng Zhang, Wei Wu, and Weiran Xu. ConSERT: A contrastive framework for self-supervised sentence representation transfer. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 5065–5075, 2021. 3

[44] Yung-Sung Chuang, Rumen Dangovski, Hongyin Luo, Yang Zhang, Shiyu Chang, Marin Soljacic, Shang-Wen Li, Wen-tau Yih, Yoon Kim, and James Glass. DiffCSE: Difference-based contrastive learning for sentence embeddings. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022. 3

[45] Xing Wu, Chaochen Gao, Liangjun Zang, Jizhong Han, Zhongyuan Wang, and Songlin Hu. ESimCSE: Enhanced sample building method for contrastive learning of unsupervised sentence embedding. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, pages 3898–3907, 2022. 3

[46] Amy X. Lu, H. Zhang, Marzyeh Ghassemi, and Alan M. Moses. Self-supervised contrastive learning of protein representations by mutual information maximization. *bioRxiv*, 2020. 3

[47] Michael Heinzinger, Maria Littmann, Ian Sillitoe, Nicola Bordin, Christine Orengo, and Burkhard Rost. Contrastive learning on protein embeddings enlightens midnight zone. *NAR Genomics and Bioinformatics*, 4(2), 2022. 3

[48] Yang Li, Guanyu Qiao, Xin Gao, and Guohua Wang. Supervised graph co-contrastive learning for drug–target interaction prediction. *Bioinformatics*, 38(10):2847–2854, 2022. 3

[49] Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021. 3

[50] Shentong Mo, Xi Fu, Chenyang Hong, Yizhen Chen, Yuxuan Zheng, Xiangru Tang, Zhiqiang Shen, Eric P. Xing, and Yanyan Lan. Multi-modal self-supervised pre-training for regulatory genome across cell types. *arXiv preprint arXiv:2110.05231*, 2021. 3

[51] Weizhi An, Yuzhi Guo, Yatao Bian, Hehuan Ma, Jinyu Yang, Chunyuan Li, and Junzhou Huang. Modna: Motif-oriented pre-training for dna language model. In *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2022. 3

[52] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 3

[53] Yue Cao, Zhenda Xie, Bin Liu, Yutong Lin, Zheng Zhang, and Han Hu. Parametric instance classification for unsupervised visual feature learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 15614–15624, 2020. 3

[54] Hu Qianjiang, Wang Xiao, Hu Wei, and Qi Guo-Jun. AdCo: adversarial contrast for efficient learning of unsupervised representations from self-trained negative adversaries. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1074–1083, 2021. 3

[55] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 3

[56] Xudong Wang, Ziwei Liu, and Stella X Yu. CLD: unsupervised feature learning by cross-level instance-group discrimination. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3

[57] Bram van Berlo, Aaqib Saeed, and Tanir Ozcelebi. Towards federated unsupervised representation learning. *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, 2020. 3

[58] Fengda Zhang, Kun Kuang, Zhaoyang You, Tao Shen, Jun Xiao, Yin Zhang, Chao Wu, Yueting Zhuang, and Xiaolin Li. Federated unsupervised representation learning. *arXiv preprint arXiv:2010.08982*, 2020. 3

[59] Grace XY Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8(1):1–12, 2017. 6

[60] Hubert Lawrence and Arabie Phipps. Comparing partitions. *Journal of Classification*, 2:193–218, 1985. 6

[61] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002. 6

[62] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. 6

[63] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019. 7