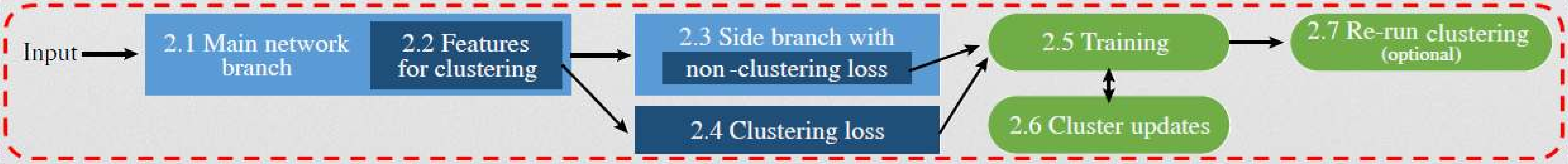
A Survey of Deep Clustering Algorithms

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

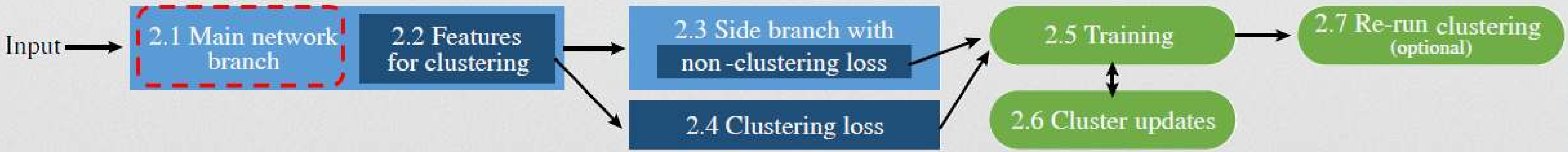
1. Clustering with Deep Learning: Taxonomy and New Methods

Clustering with Deep Learning: Taxonomy and New Methods



- Neural network training procedure, consisting of:
  - Main neural network branch and its usage
    - Architecture of main neural network branch, described in Section 2.1
    - Set of deep features used for clustering, described in Section 2.2
  - Neural network losses:
    - Non-clustering loss, described in Section 2.3
    - Clustering loss, described in Section 2.4
    - Method to combine the two losses, described in Section 2.5
  - Cluster updates, described in Section 2.6
- (Optional) Re-run clustering after network training, described in Section 2.7

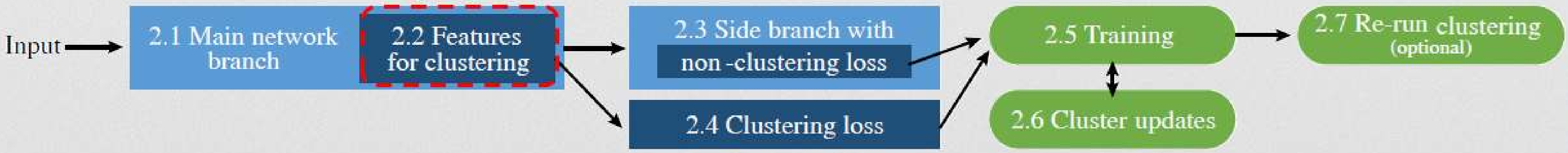
Clustering with Deep Learning: Taxonomy and New Methods



- **Architecture of Main Neural Network Branch:**
  - Multilayer Perceptron (MLP)
  - Convolutional Neural Network (CNN)
  - Deep Belief Network (DBN)
  - Generative Adversarial Network (GAN)
  - Variational Autoencoder (VAE)

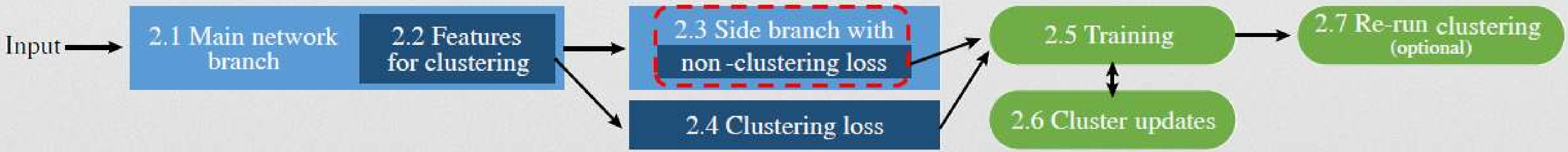
注：该框表示本文提出的方法所用到的技术(下同)

Clustering with Deep Learning: Taxonomy and New Methods



- **Set of Deep Features Used for Clustering:**
  - ➢ One layer
    - Refers to the case where only one layer of the network is used which is beneficial because of its low dimensionality. In most cases <u>the output of the last layer is used</u>.
  - ➢ Several layers
    - Refers to the case where the representation is <u>a combination of the outputs of several layers</u>. Thus, the representation is richer and allows the embedded space to represent more complex semantic representations, which might enhance the separation process and help in the similarity computation (Saito and Tan, 2017).

Clustering with Deep Learning: Taxonomy and New Methods



- **Non-Clustering Loss:**
  - No non-clustering loss

    目标函数只有聚类，没有非聚类的损失函数
  - Autoencoder reconstruction loss

$$L = d_{\mathrm{AE}}(x_i, f(x_i)) = \sum_i \|x_i - f(x_i)\|^2$$

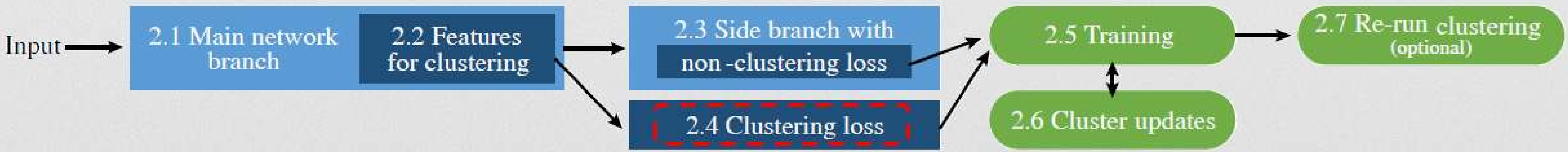
  - Self-augmentation loss

$$L = -\frac{1}{N}\sum_N s(f(x), f(T(x)))$$

where x is the original sample, T is the augmentation function, $f(x)$ is the representation generated by the model, and s is some measure of similarity (for example cross-entropy if f has a softmax nonlinearity).

  - Other tasks

Clustering with Deep Learning: Taxonomy and New Methods



- **Clustering Loss:**
  - ➢ No clustering loss
    - 目标函数没有聚类的损失函数，深度学习提取的特征作为聚类的输入数据
  - ➢ k-Means loss

$$L(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} s_{ik} \|z_i - \mu_k\|^2$$

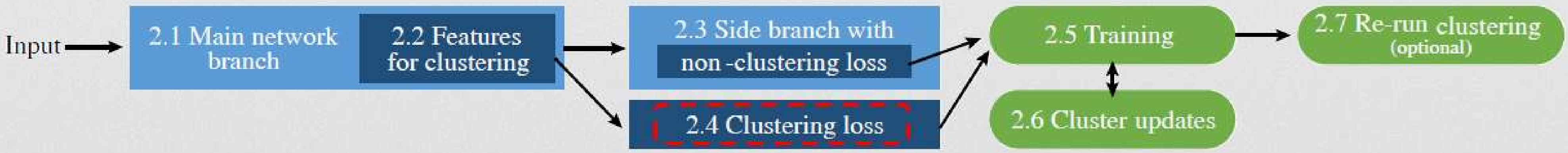
  - ➢ Cluster assignment hardening

$$L = \mathrm{KL}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \qquad q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2/\nu)^{-\frac{\nu+1}{2}}}{\sum_{j'} (1 + \|z_i - \mu_{j'}\|^2/\nu)^{-\frac{\nu+1}{2}}} \qquad p_{ij} = \frac{q_{ij}^2/\Sigma_i q_{ij}}{\Sigma_{j'}(q_{ij'}^2/\Sigma_i q_{ij'})}$$

  - ➢ Balanced assignments loss

$$L_{ba} = \mathrm{KL}(G\|U) \qquad g_k = P(y=k) = \frac{1}{N} \sum_i q_{ik}$$

Clustering with Deep Learning: Taxonomy and New Methods



- **Clustering Loss:**
  - Locality-preserving loss

$$L_{lp} = \sum_{i} \sum_{j \in N_k(i)} s(x_i, x_j) \|z_i - z_j\|^2$$

x:数据，z:中间那一层

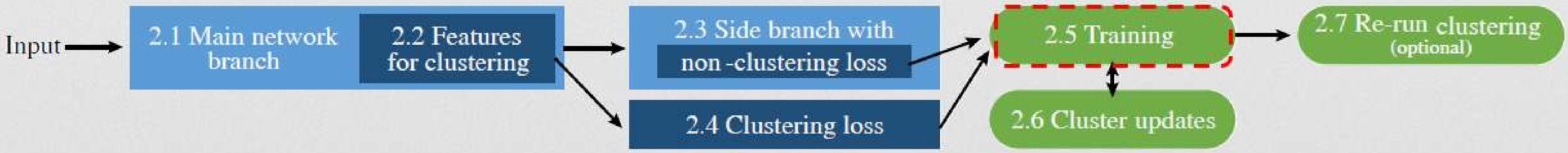
  - Group sparsity loss

$$L_{gs} = \sum_{i=1}^{N} \sum_{g=1}^{G} \lambda_g \|\phi^g(x_i)\|, \quad \lambda_g = \lambda\sqrt{n_g},$$

G:聚类数

  - Cluster classification loss
    - 聚类更新获得的类标签作为分类损失函数的伪类标签
  - Agglomerative clustering loss(层次聚类的一种)
    - 按照一定规则，将最满足规则条件的两个类进行合并，直到满足既定条件停止迭代

Clustering with Deep Learning: Taxonomy and New Methods



● Method to Combine the Losses:

$$L(\theta) = \alpha L_c(\theta) + (1 - \alpha) L_n(\theta)$$

➢ 参数α的确定方法： Clustering Loss Network Loss

- Pre-training, fine-tuning(分两个阶段，只取0,1)

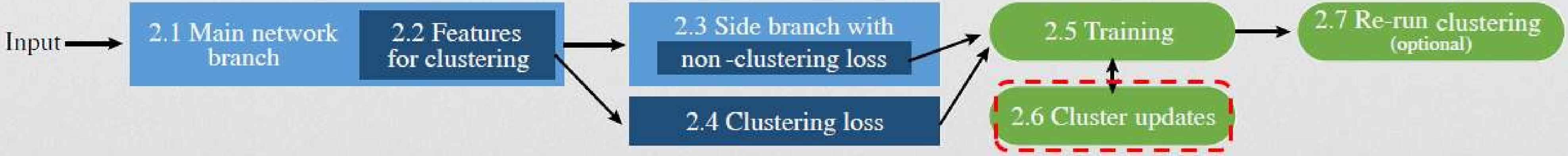
First, α=0, i.e. the network is trained using the non-clustering loss only. Subsequently, α=1, i.e. the non-clustering network branches (e.g. autoencoder’s decoder) are removed and the clustering loss is used to train (fine-tune) the obtained network. 例如：DEC

- Joint training(固定α为[0,1]某一值)

0<α<1, for example α=0.5, i.e. the network training is affected by both loss functions.

- Variable schedule(随迭代次数的变化而变化)

α is varied during the training dependent on a chosen schedule. For instance, start with a low value for and gradually increase it in every phase of the training.

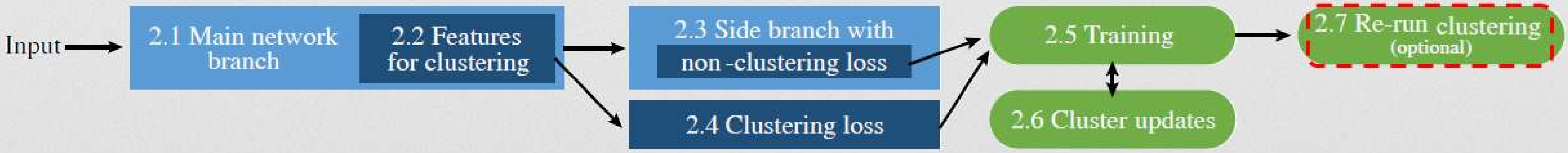


- **Method to Combine the Losses:**
  - Jointly updated with the network model

    簇划分矩阵是概率分布，软划分，将其作为参数用BP更新。
  - Alternatingly updated with the network model

    簇划分矩阵是硬划分，主要依赖于以下两个因素：
    - 迭代次数
    - 更新频率

Clustering with Deep Learning: Taxonomy and New Methods



- **After Network Training:**
  - 重新训练聚类算法的原因
    - Clustering a similar dataset

      To reuse the learned features representation mapping on another dataset which is similar to the one that has been used but has different data.
    - Obtaining better results

      It is possible that the results of clustering after the training are better than the ones obtained during the learning procedure.

2. A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture

A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture

- Neural Network Architecture for Deep Clustering:
  - Feedforward full-connected neural network
  - Feedforward convolutional neural network
  - Deep belief network
  - Autoencder
  - GAN & VAE
- Loss Function Related to Clustering:
  - Principal clustering loss(在学习特征表示的过程中同时聚类)

    These loss functions contain the cluster centroids and cluster assignments of samples. e.g. k-means loss, cluster assignment hardening loss, agglomerative clustering loss and nonparametric maximum margin clustering.)
  - Auxiliary clustering loss(学习特征表示完之后聚类)

    E.g. locality-preserving loss, group sparsity loss, sparse and subspace clustering loss.
- Performance Evaluation Metrics for Deep Clustering:

$$ACC = \max_m \frac{\sum_{i=1}^{n} 1\{y_i = m(c_i)\}}{n} \qquad NMI(Y, C) = \frac{I(Y, C)}{\frac{1}{2}[H(Y) + H(C)]}$$

TABLE 1. Comparison of different categories of deep clustering algorithms.

Categories	L_c	L_n	**Description**	**Advantages**	**Disadvantages**	**Computational Complexity**
AE-based DC	Yes	Yes (AE loss)	Joint optimize an AE and clustering parameters	1) Not obtain trivial solutions 2) Easy to implement	1) Introduce a hyper-parameter to balance the two losses 2) Limited network depth	Clustering loss specific
CDNN-based DC	Yes	No	Optimize the network only by clustering loss	1) Simple and graceful objective 2) Extended to large-scale tasks	1) Have the risk of obtaining corrupted feature space 2) Require well-designed clustering loss	Clustering loss specific
VAE-based DC	Yes	Yes	Impose a GMM priori on VAE	1) Capable to generate samples 2) Decent theoretical guarantee	High-computional complexity	High
GAN-based DC	Yes	Yes	Impose a multi-class priori on GAN	1) Capable to generate samples 2) Flexible	1) Hard to converge 2) Mode collapse	High

- Taxonomy of Deep Clustering:

$$L = \lambda L_n + (1-\lambda) L_c$$

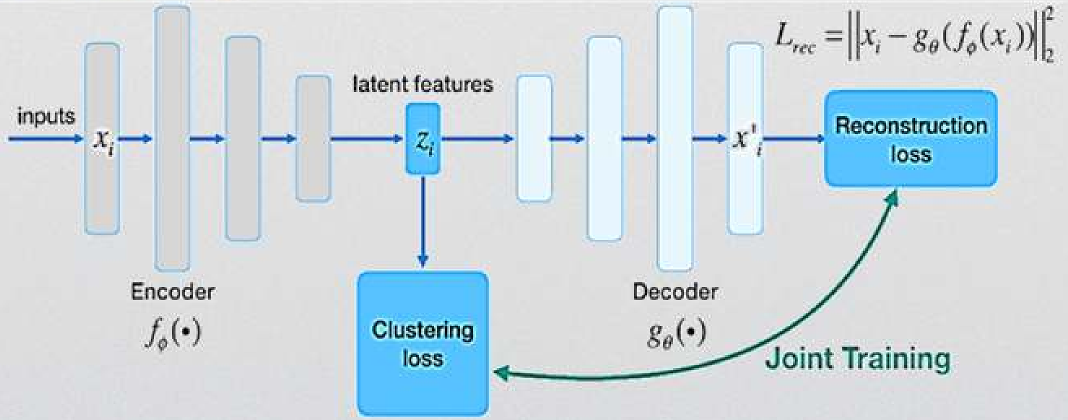
Network Loss   Clustering Loss

- AE-Based Deep Clustering

$$\min_{\phi,\theta} L_{rec} = \min \frac{1}{n} \sum_{i=1}^{n} \| x_i - g_\theta(f_\phi(x_i)) \|^2 \qquad L = \lambda L_{rec} + (1-\lambda) L_c$$

- 可改进的地方
  - Architecture(e.g., convolutional autoencoder(CAE))
    - 为了处理空间不变性的数据，如图像数据，可以增加卷积、池化层
  - Robustness(e.g., denoising autoencoder)
    - 为了避免过拟合，在输入层加噪声
  - Restrictions on latent features(e.g., sparse autoencoder)
    - 为了约束隐层特征，比如隐层维度低于输入数据的维度，约束隐层为稀疏矩阵
  - Reconstruction loss
    - 为了重构输入与输出数据之间的差异性，也可以重构所有层的损失函数

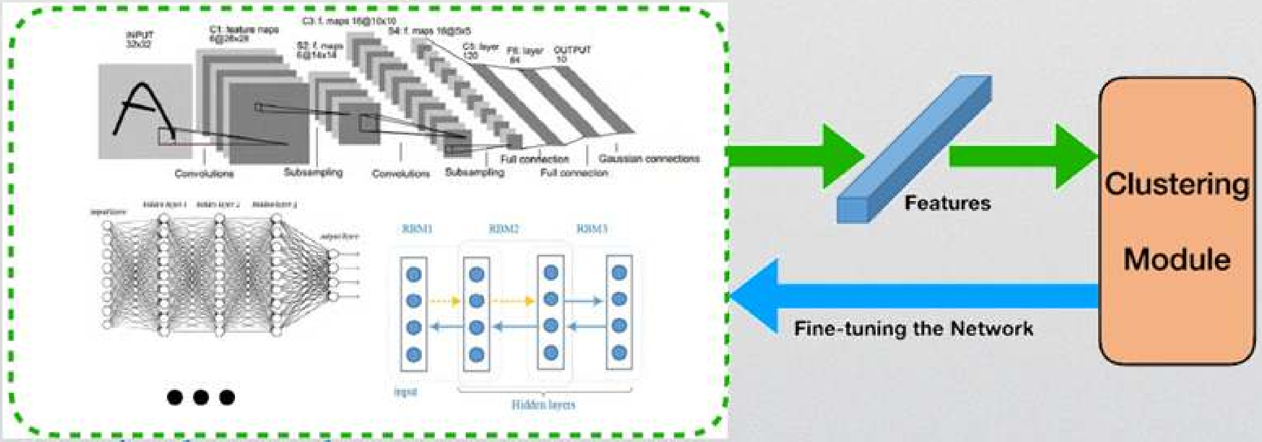
- AE-Based Deep Clustering



  - Deep Clustering Network (DCN)
  - Deep Embedding Network (DEN)
  - Deep Subspace Clustering Networks (DSC-Nets)
  - Deep Multi-Manifold Clustering (DMC)
  - Deep Embedded Regularized Clustering (DEPICT)
  - Deep Continuous Clustering (DCC)

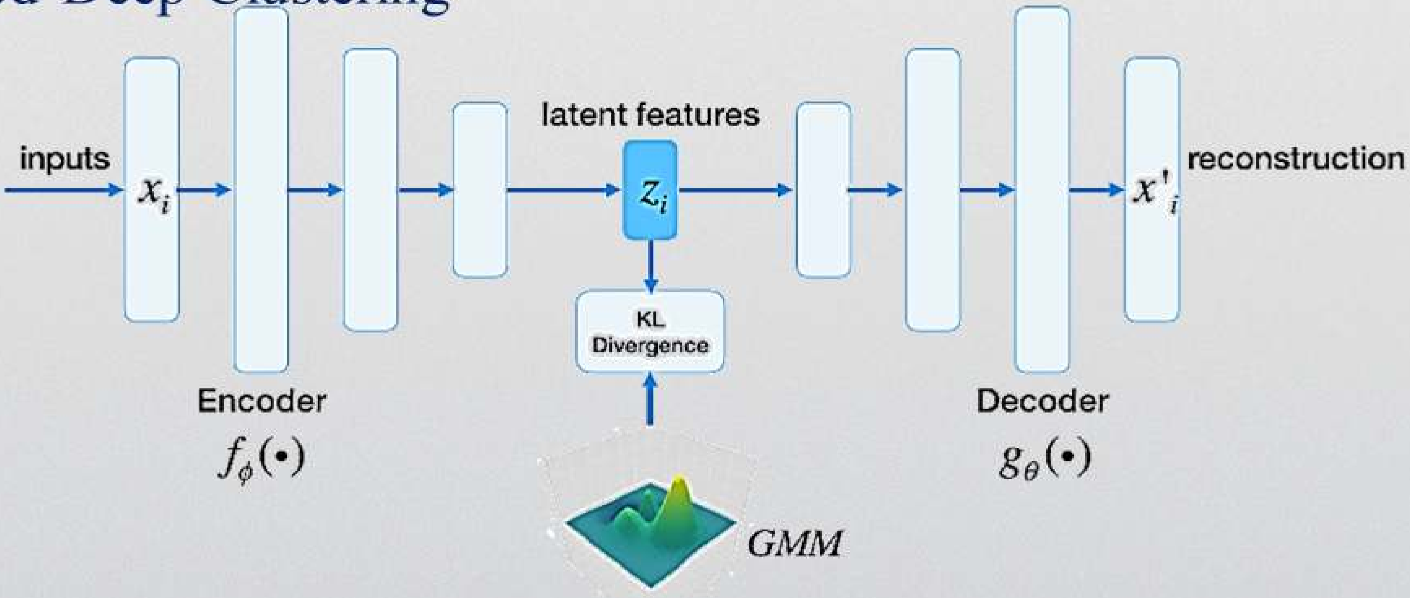
CDNN-Based Deep Clustering

$$L = L_C$$



- Unsupervised pre-trained network
  - Deep Nonparametric Clustering (DNC)
  - Deep Embedded Clustering (DEC)
  - Discriminatively Boosted Clustering (DBC)
- Supervised pre-trained network
  - Clustering Convolutional Neural Network (CCNN)
- Non-pre-trained network
  - Information Maximizing Self-Augmented Training(IMSAT)
  - Joint Unsupervised Learning (JULE)
  - Deep Adaptive Image Clustering (DAC)

- VAE-Based Deep Clustering



$$L(\theta, \phi; X) = \sum_{i}^{N}(-D_{KL}(q_\phi(z|x^{(i)}) \parallel p(z)) + \mathbb{E}_{q_\phi(z|x^{(i)})}[logp_\theta(x^{(i)}|z)])$$

- Variational Deep Embedding (VaDE)
- Gaussian Mixture VAE (GMVAE)

- GAN-Based Deep Clustering

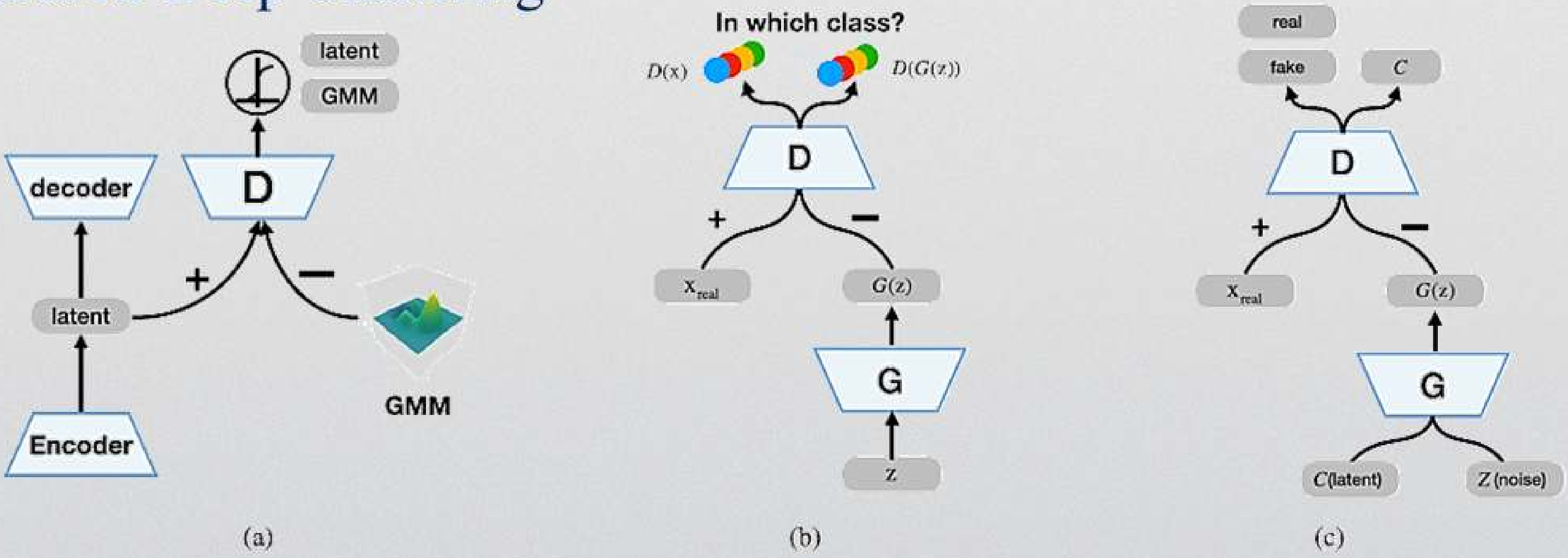


FIGURE 4. GAN-based deep clustering. (a) DAC. (b) CatGAN. (c) InfoGAN.

$$\min_{G}\max_{D}\mathbb{E}_{x\sim p_{data}}[\log D(x)]+\mathbb{E}_{z\sim p(z)}[\log(1-D(G(z)))]$$

- Deep Adversarial Clustering (DAC)
- Categorial Generative Adversarial Network(CatGAN)
- Information Maximizing Generative Adversarial Network (InfoGAN)

Future Opportunities of Deep Clustering

Theoretical exploration

Although jointly optimizing networks and clustering models significantly boost the clustering performance, there is no theoretical analysis explaining why it works and how to further improve the performance.

Other network architectures

Existing deep clustering algorithms mostly focus on image datasets, while few attempts have been made on sequential data, e.g., documents.

Tricks in deep learning

It is viable to introduce some tricks or techniques used in supervised deep learning to deep clustering, e.g. data augmentation and specific regularizations. A concrete example is augmenting data with noise to improve the robustness of clustering methods.

Other clustering tasks

Combining deep neural networks with diverse clustering tasks, e.g. multi-task clustering, self-taught clustering (transfer clustering) is another interesting research direction.

Comparison of algorithms based on network architecture and loss function.

Categories	Algorithms	Network Architecture	Network loss	Clustering loss	
				Principal	Auxiliary
AE	DCN	AE	reconstruction loss	k-means loss	N
	DEN	AE	reconstruction loss	N	1) locality-preserving constraint 2) group sparsity constraint
	DSC-Nets	CAE	reconstruction loss	N	self-expressiveness term
	DMC	AE	reconstruction loss	proximity penalty term	locality-preserving loss
	DEPICT	CAE (Denoising)	reconstruction loss	unsupervised cross entropy loss	N
	DCC	AE/CAE	reconstruction loss	robust continuous clustering loss	N
CDNN	DNC	RBM	N	nonparametric maximum margin clustering loss	N
	DEC	FCN	N	cluster assignment hardening loss	N
	DBC	CNN	N	cluster assignment hardening loss	N
	CCNN	CNN	N	k-means	N
	IMSAT	FCN	N	1) regularized information maximization, 2) self-augmented training loss	N
	JULE	CNN	N	agglomerative clustering	N
	DAC[1]	CNN	N	pairwise-classification loss	N
VAE	VaDE	VAE	variational lower bound on the marginal likelihood, with a GMM priori		
	GMVAE	VAE	variational lower bound on the marginal likelihood, with a GMM priori		
GAN	DAC[2]	Adversarial autoencoder	reconstruction loss	1) GMM likelihood, 2) adversarial objective	N
	CatGAN	GAN	adversarial objective with a multi-classes priori		
	InfoGAN	GAN	adversarial objective with a multi-classes priori		

[1] Deep Adaptive Clustering
[2] Deep Adversarial Clustering

Main contributions of the representative algorithms.

Categories	Algorithms	Main contributions to clustering
AE	DCN	perform k-means clustering and feature learning simultaneously, simple but effective
	DEN	learn a clustering-friendly representation
	DSC-Nets	improve the classical subspace clustering by AE
	DMC	improve the classical multi-manifold clustering by AE
	DEPICT	computational efficient, robust, perform well on image datasets
	DCC	avoid alternative optimization, require no prior knowledge of cluster number
CDNN	DNC	improve the classical NMMC clustering by DBN
	DEC	the first well-known deep clustering method, making this field popular
	DBC	improve DEC using CNN
	CCNN	computational efficient, deal with large-scale image datasets
	IMSAT	introduce self-augment training to deep clustering
	JULE	perform well on image datasets, but have high computational and memory cost
	DAC	well-designed clustering loss, achieve the-state-of-art performance on several datasets
VAE	VaDE	combine VAE with clustering
	GMVAE	combine VAE with clustering
GAN	DAC	combine AAE with clusteirng
	CatGAN	combine GAN with clustering
	InfoGAN	learn disentangled representations

3. 参考文献

[1] Aljalbout E, Golkov V, Siddiqui Y, et al. Clustering with deep learning: Taxonomy and new methods[J]. arXiv preprint arXiv:1801.07648, 2018.

[2] Min E, Guo X, Liu Q, et al. A survey of clustering with deep learning: From the perspective of network architecture[J]. IEEE Access, 2018, 6: 39501-39514.

[3] Deep Clustering: methods and implements-Github 深度聚类会议论文汇总