大连理工大学本科毕业设计（论文）

基于联邦学习的多模态聚类方法

Multimodal clustering method based on federated learning

学 院（系）：电子信息与电气工程学部

专　　业：计算机科学与技术

学 生 姓 名：武振威

学　　号：201971028

指 导 教 师：孙景昊

评 阅 教 师：刘婵娟

完 成 日 期：2023.5.28

大连理工大学

Dalian University of Technology

原创性声明

本人郑重声明：本人所呈交的毕业设计（论文），是在指导老师的指导下独立进行研究所取得的成果。毕业设计（论文）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究成果做出重要贡献的个人和集体，均已在文中以明确方式标明。

本声明的法律责任由本人承担。

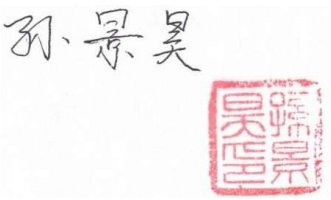
作者签名：武振威　　日　期：2023 年 6 月 9 日

关于使用授权的声明

本人在指导老师指导下所完成的毕业设计（论文）及相关的资料（包括图纸、试验记录、原始数据、实物照片、图片、录音带、设计手稿等），知识产权归属大连理工大学。本人完全了解大连理工大学有关保存、使用毕业设计（论文）的规定，本人授权大连理工大学可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业设计（论文）。如果发表相关成果，一定征得指导教师同意，且第一署名单位为大连理工大学。本人离校后使用毕业毕业设计（论文）或与该论文直接相关的学术论文或成果时，第一署名单位仍然为大连理工大学。

论文作者签名：武振威　　日　期：2023 年 6 月 9 日

指导老师签名：　　日　期：2023 年 6 月 9 日



摘　要

多模态数据的复杂性、海量性使得难以对其进行精准标注，这使得以多模态聚类为代表的无监督多模态学习成为流行的多模态数据处理手段。然而，多模态数据聚类面临诸多困难。一方面，出于隐私保护的考量，不同数据源难以直接共享数据，而单一数据源的数据在种类和数量上有限；另一方面，由于传感器故障等因素，模态缺失问题较为普遍。针对上述难题，本文对基于联邦学习的缺失多模态聚类技术进行研究，主要内容如下：

(1)针对多源缺失多模态数据联合学习问题，本文首先将融合了自编码器和循环生成对抗网络（Cycle Generative Adversarial Network, CycleGAN）的缺失多模态生成模型 VIGAN 拓展到联邦学习场景，利用联邦平均算法(Federated Averaging Algorithm, FedAvg)构建联邦训练方案 Fed-VIGAN，以克服单一数据源因数据不足导致的缺失模态补全难题。

(2)本文基于 PyTorch 实现了所提出的 Fed-VIGAN，在 MNIST 数据集上进行模型训练与测试，并设计一系列的测试对比实验。结果表明 Fed-VIGAN 能够在限制客户端之间数据交互、保护隐私的前提下训练出与集中式 VIGAN 性能相近的全局模型，同时验证了所提出的联邦缺失多模态的聚类有效性。

本文的研究内容能够有效解决多源、多模态、缺失场景下的数据处理难题，对于推进多模态数据的应用、保护数据隐私等方面具有重要意义。

关键词：联邦学习；多模态数据缺失；多模态聚类；生成对抗网络；多模态自编码器

Multimodal clustering method based on federated learning

Abstract

The complexity and mass of multimodal data make it difficult to accurately label it, which makes unsupervised multimodal learning represented by multimodal clustering become a popular multimodal data processing method. However, multimodal data clustering faces many difficulties. On the one hand, due to privacy protection considerations, it is difficult for different data sources to share data directly, and the type and quantity of data from a single data source is limited. On the other hand, due to sensor failure and other factors, the problem of mode loss is more common. To solve the above problems, this paper studies the missing multimodal clustering technology based on federated learning. The main contents are as follows:

(1)Aiming at the problem of multi-source missing multi-modal data joint learning, this paper first extends VIGAN, a missing multi-modal generation model that integrates autoencoder and Cycle Generative Adversarial Network (CycleGAN), to federated learning scenarios. Using Federated Averaging Algorithm (FedAvg), a federal training scheme Fed-VIGAN was constructed to overcome the problem of missing mode completion caused by insufficient data of a single data source.

(2)The proposed Fed-VIGAN is implemented based on PyTorch, and the model is trained and tested on the MNIST dataset, and a series of test and comparison experiments are designed. The results show that Fed-VIGAN can train a global model with similar performance to centralized VIGAN under the premise of limiting the data interaction between clients and protecting privacy, and verify the clustering effectiveness of the proposed federated missing multimode.

The work in this thesis can well solve the problem of multi-source missing-modal data processing and could contribute to the application of multi-modal data as well as privacy preservation.

Key Words： Federated learning; Multimodal data missing; Multimodal clustering; Generate adversarial network; Multimode autoencoder

目　录

引　言

如今，随着网络边缘设备的普及以及先进的数据采集技术的发展，目前越来越多的数据以多模态的形式存在，例如：图像数据不仅可以通过它的视觉特征来表示，也可以通过描述他的文本的数据来表示；一个人的身体状态不仅可以由传感器生成的脑电图表示，也可以通过心电图来表示。与单一模态数据相比而言，多模态数据具有互补性和一致性的特点。其中，互补性指的是多模态数据中的不同模态之间相互补充的性质，通过结合多种模态的信息，可以得到更全面、准确的描述和认知；一致性则指的是多模态数据中的不同模态之间具有一致的特征和关联，因此多模态数据通常可以通过一些隐含的关联将不同模态的信息联系起来进行训练，已达到更好的学习效果。然而，多模态数据多源、海量、异构的特性使得难以获取到大量精确标签，这使得以多模态聚类为代表的无监督多模态学习技术成为一种流行的多模态数据处理手段，并得到了国内外学者的广泛关注。

目前，大多数多模态学习方法大多针对不同模态数据集中训练的场景，并且集中训练对多模态数据的质量和数量都有一定的要求，但是在实际情况中，单一机构所拥有的数据往往规模有限、数据种类不够丰富或者数据模态有所缺失，这使得单一数据持有者难以单独训练一个功能完善、效果良好的模型。但出于数据隐私与数据价值的考虑，不同数据持有方往往难以实现对数据的直接共享。近年来，数据泄露事件频发，如某家智能家居公司所拥有的超过 20 亿条的物联网日志被泄露、美国医疗机构 Medical Collection Agency 设计 2000 万客户的 2000 万条付款记录被泄露间接导致该机构的破产，这使得隐私问题得到了越来越多的关注。在国家层面也对隐私问题相继颁布相应措施，如中国颁布的《个人隐私数据保护法》、《中华人民共和国网络安全法》以及欧盟的通用数据保护条例（General Data Protection Regulation，GDPR）。因此如何在保护隐私的前提下对训练多模态数据，成为一个亟待解决的问题。

联邦学习作为一种新兴的机器学习框架，在构建个性化模型方面表现出了强大的潜力不同于传统的集中式训练方法，联邦学习允许在分布式设备上进行本地训练，将模型更新的信息聚合在一起以形成全局模型，而无需在个人设备上共享原始数据。这种分布式训练的方式使得联邦学习成为一个可以满足多数据源协同学习的需求的理想解决方案。由于数据通常分散在不同的设备或数据中心中，联邦学习允许这些设备通过本地训练来改进模型，并将有关模型的更新信息发送回中央服务器进行中央模型聚合从而训练出基于本地数据总和的中央模型。最重要的是，由于数据留在本地设备上进行训练，并且仅发送模型更新的梯度或聚合参数，联邦学习可以避免直接暴露个人隐私信息，具有

显著的隐私保护优势。综合考虑联邦学习的特点以及多模态学习的需求，可以将联邦学习应用于多模态学习以实现具有隐私保护的多模态数据协同处理方案。目前有许多学者已经进行该方面的研究并且在不同的领域中均有较为良好的表现。比如 Oyomno W 等人[1]利用联邦多视图矩阵分解方法为用户提供个性化推荐，并且提出了联邦学习模式下冷启动预测机制，而在相同的领域，Huang Mingkai 等人[2]以联邦学习为基础框架基于深度结构化语义模型（Deep Structured Sematic Models，DSSM），提出了一个通用的基于内容的联邦多视图框架称为 FL-MV-DSSM（Federated Multiview DDSM）， 它从多个数据源学习联邦模型来捕获更丰富的用户级特征，能够显著提高推荐性能。Che Sicong 等人[3]则将联邦多模态学习应用于医疗领域，他们提出一个通用的联邦多视图学习框架，该框架具有两种模型来处理于不同的分布类型的多试图数据，从而来整合分析多个医疗站点所持有的隐私数据，最终效果优于站点局部训练。物物互联，联邦多模态学习在物联网领域也有所应用，Zhao Yuchen 等人[4]提出一种多模态联邦平均算法，通过聚合在物联网环境、不同数据模式（兼顾单模态与多模态）下训练的本地自编码器来全局聚合形成一个全局自编码器，大大提高了分类性能。

以上联邦多模态学习方法虽然取得了较好的效果，但是这些模型均假设多模态数据的完备性，而在实际应用中，在数据的采集、传输与存储过程中，很容易因为传感器故障、信道噪声或者存储不当导致数据缺失问题（如图 1 所示）。处理缺失多模态数据可简单分为两种，即删除包含缺失值的全部缺失值（删除整个缺失样本）和对缺失值进行补全，即完善样本。其中前者不仅会再次减少数据集的可用数据规模，同时也会因此丢失整一个模态数据而降低模型的泛化能力。因此为了训练性能表现更好的模型就要求对多模态数据进行补全以提供更为完备的数据集。目前，已经有大量关于缺失多模态数据补全方法。其中一种经典的解决方法是利用生成模型实现缺失数据的补全扩充，典型技术是生成对抗网络（Generative Adversarial Network, GAN）。然而，高缺失率或数据采集场景的有限性，单一机构所持有的数据在分布上通常不满足独立同分布特性，这极大的限制了 GAN 等生成模型的生成效果。针对该问题，一些学者利用联邦学习实现数据间的扩充，同时支持隐私保护。

针对多方协同的联邦数据补全问题，Zhou 等人[5]考虑单个机构采集的数据不满足独立同分布要求，引入联邦学习实现具有隐私保护的模型协同训练，其主要思路是利用 FedAVG 方法在多个机构间建立横向联邦学习机制，完成对本地数据的扩充。每个客户端基于本地现有数据和表示缺失数据的 mask 更新服务器的全局 GAN 副本，并将修改后的参数传回服务器，服务器将平均参数以更新全局 GAN，最终即可利用全局 GAN 补全数据，提供训练过程中的数据随机缺失数据机制可以使得数据补全的效果更

佳。Yang Bing 等[6]人提出了一种用于缺失交通数据输入的时空可学习双向生成对抗网络(ST-LBAGAN)，改名模型采用具有注意力机制的编码器和解码器结构提取特征，并以历史观测、不完整数据和被屏蔽图像作为生成器的输入，利用二值分类作为鉴别器的输出来获得缺失数据的估计。但是以上方法均针对的是单模态数据缺失问题而非多模态数据数据缺失问题。

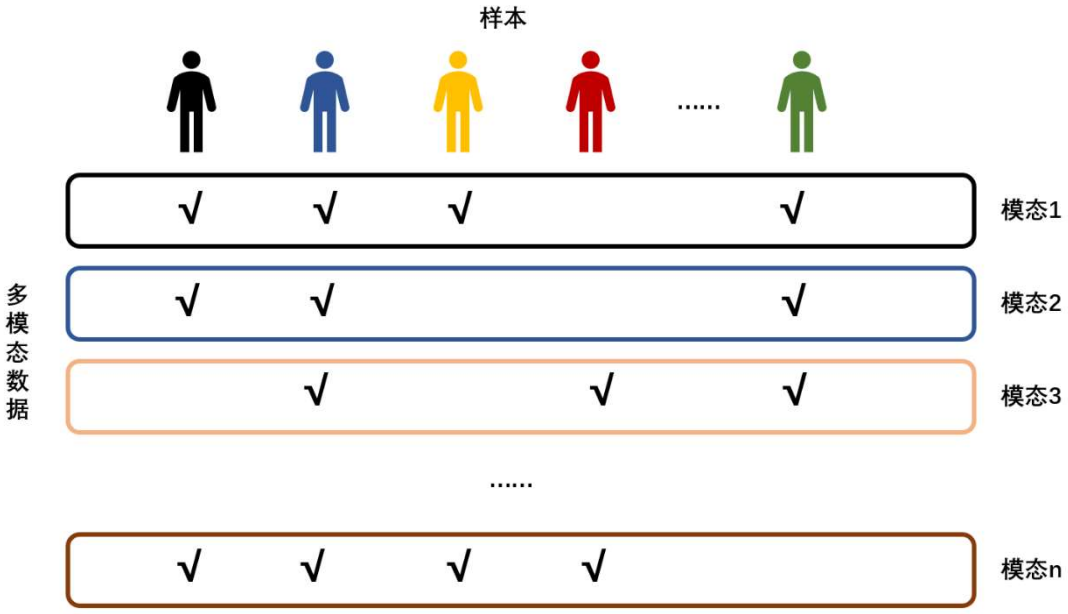


图 1　多模态数据缺失图例，其中只有第二个样本模态齐全

此外，一些工作专注于对本地样本的扩充方法进行研究，其主要思路是利用联邦 GAN 聚合全局样本信息以构建样本生成器扩充本地样本。例如：考虑到联邦学习多个客户端之间数据的分布可能是非同分布的，而需要基于全局的样本分布来补全本地的数据，因此 Zhang Yikai 等人[7]提出一种称为 UA-GAN 的通用聚合方法，该方法可用聚合多个客户端的数据，从而学习得到一个可用生成混合分布数据的生成器。Xie Guoyang 等人[8]基于医学图像缺失，提出了 FedMed-GAN 用于生成脑图像，该模型在服务器维护一个全局的生成器，而在每个客户端维护一个局部鉴别器，通过不断地训练在服务器和客户端之间传递参数，从而训练出一个全局生成器。同时该模型通过联邦平均算法处理梯度惩罚，然后利用差分隐私梯度下降来正则化训练动态从而一定程度上减少 GAN 的模式崩溃。此外，一些学者提出，仅仅避免在客户端和服务器或者客户端和客户端之间直接分享数据不能有效防止攻击者提取用户隐私信息，于是他们在引入差分隐私思想，利用 DP-SGD 更新模型以构建支持差分隐私的联邦 GAN 方法[9]。

本文基于[21]中提出的一种 GAN 和自动编码器的复合方法，并利用联邦学习基于全局数据集提高生成器的生成效果来构建本地的多模态数据补全模型。该方法可以通过一个两步的训练过程来训练模型从而生成缺失的模态。本文的工作创新点以及贡献如下：

a) 本文针对多源缺失多模态数据联合学习，提出了基于缺失多模态生成模型 VIGAN 的联邦训练方案 Fed-VIGAN。该方案建立在联邦环境下，对保护数据持有者的隐私具有重要意义。

b) 所提出的 Fed-VIGAN 能够综合学习全局数据的数据分布，以用于补全本地的多模态数据，提高补全缺失多模态数据的效果。

c) 本文基于 PyTorch 框架实现了所提出的联邦缺失数据生成模型 Fed-VIGAN 以及联邦聚类算法，在 MNIST 手写数字集上进行实验，进一步验证了所提出的联邦缺失多模态聚类方法的有效性。

本文剩余部分内容如下：第一章进行相关工作的介绍，包括 GAN、自编码器以及联邦学习等工作的背景以及概念；第二章介绍所提出的 Fed-VIGAN 模型；第三章基于 MNIST 手写数字数据集对所提出模型的生成性进行实验，并在第四章利用 K-means 聚类算法对其聚类有效性进行分析测试；最后，在结论中对本文工作进行了总结并对下一步工作进行了展望。

1 相关原理介绍

本文所研究的联邦缺失多模态聚类方法主要基于联邦学习、生成对抗网络等模型进行设计。为更好的对本文工作进行介绍，本章内容对多模态数据缺失的数据处理方法、生成模型等相关技术和工作进行了整理，并分析了这些工作在处理缺失多模态数据时的不足。

1.1 矩阵补全算法

矩阵补全算法是一种填充缺失数据的算法，它的重点是在一定条件下对部分观测数据所形成的数据矩阵的缺失项进行补全。具体来说，在矩阵补全算法中，一般会将数据集表示为一个矩阵，其中每一列代表一个样本，而列中元素就是样本的特征，每一行对应相同的特征。常用的矩阵补全算法可分为两类，即基于矩阵近似的方法和基于矩阵分解的方法。

在基于矩阵分解的方法中，有一种称为非负矩阵分解（Non-negative Matrix Factorization，NMF）[11]。该方法中将非负数据集表示为矩阵$X \in R^{d\times n}$，其中d表示特征维度，n表示样本数，然后试图将该矩阵分解成两个更低维度的非负矩阵的乘积，即：

$$X = UV^T \tag{1.1}$$

其中$U \in R^{d\times c} \geq 0, V \in R^{n\times c} \geq 0, c \leq \min(n, d)$。

同时由于X是模态数据有缺失的数据集，其第i个样本的第j个特征在数据集中的表示x_{ji}可能为 0 或者 NAN，无法计算出确切的低维度矩阵U, V，因此需要采用一种近似得解法，即使得UV^T与X最为接近：

$$X \approx \bar{X} = UV^T \tag{1.2}$$

因此目标函数为：

$$argmin_{U,V} \ ||X - UV^T||_F^2 \tag{1.3}$$

最终通过调整U, V的参数和维度最小化目标函数即可获得一个近似矩阵$\bar{X}$用于填充缺失值。但是 NMF 分解所得的特征向量存在冗余，因此基于奇异值分解（Singular Value decomposition，SVD）[12]的矩阵分解方法被提出，该方法保证特征向量之间彼此正交，但是失去了矩阵非负的特点，会导致解释性变差。

矩阵近似的方法则是通过最小化原始矩阵和近似矩阵之间的近似距离来获得一个最佳的近似矩阵来填补缺失值。虽然基于矩阵补全的算法模型简单易于实现，当其运用到多模态数据补全上，则会出现以下问题：第一，由于多模态数据往往数据规模较大，

而矩阵补全算法需要多次迭代，它的算法复杂度往往是$O(N^3)$的，因此会导致计算量过于庞大；第二，由于在矩阵补全算法中，往往会把一个样本的特征表示为矩阵的一列，但是对于多模态数据这个方法有不足之处，因为直接把多个模态特征简单的连接为一个向量会无法合理利用模态与模态间的依赖关系。因此以上问题限制了基于矩阵的补全算法在多模态领域的应用。

1.2 生成对抗网络

生成模型是一种能够学习数据分布并生成新样本的算法，它们通过观察现有数据样本的统计特征来推断缺失值，并生成逼近真实数据分布的补全结果。与传统的矩阵补全方法相比，生成模型具有显著的优势。首先，生成模型能够学习数据样本之间的潜在关系和内在结构，从而能够更好地理解数据的上下文信息。这使得生成模型在补全缺失值时能够考虑到更多的相关因素，提高了补全结果的准确性，这对多模态数据补全是有利的。其次，生成模型还能够通过对数据分布的建模来生成新的合成数据样本，从而扩充数据集。这在数据稀缺或不完整的情况下尤为有用，可以增加训练数据量，提高模型的性能和鲁棒性。

生成模型中比较流行的一种是生成对抗网络（Generative Adversarial Network，GAN），模型结构如图 1.1 所示。GAN 由一个生成器（Generator）和一个鉴别器（Discriminator）组成。其中，生成器试图生成逼真的数据样本，而判别器则尝试区分生成器生成的样本和真实数据样本，GAN 就是利用零和博弈的特殊训练思想使得其在数据生成方面取得了优秀的效果。

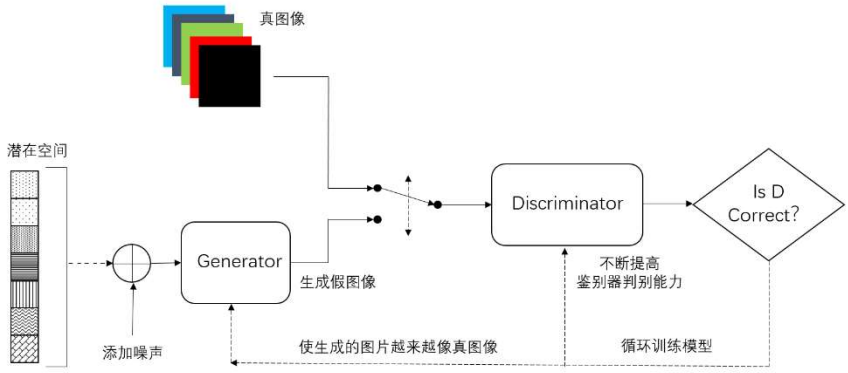


图 1.1 生成对抗网络 GAN 模型图

GAN 模型的训练采用交替迭代训练的方法。首先接受服从任意分布P_z的随机采样噪声作为生成器输入，使用生成器生成假样本$G(z)$，固定生成器的参数θ_g，通过公式

(1.4)中损失函数更新鉴别器的参数θ_d，即试图满足鉴别器对应真实样本的输出为 1 而对应生成样本的输出为 0 的目标：

$$L_D = E_{x\sim P_{data}(x)}\, log\big(D(x)\big) + E_{z\sim P_z(z)}\, log\left(1 - D\big(G(z)\big)\right) \tag{1.4}$$

然后固定鉴别器的参数θ_d，过公式(1.5)中损失函数更新鉴别生成器的参数θ_g，即试图让鉴别器将生成的图像鉴别为真实的样本，即：

$$L_G = E_{z\sim P_z(z)}\, log\left(1 - D\big(G(z)\big)\right) \tag{1.5}$$

最终 GAN 的训练将是生成器和鉴别器两个模型的最大化—最小化博弈，最终的目标函数如公式(1.6)所示：

$$min_G max_D E_{x\sim P_{data}(x)}\, log\big(D(x)\big) + E_{z\sim P_z(z)}\, log\left(1 - D\big(G(z)\big)\right) \tag{1.6}$$

GAN 的训练希望找到生成器与判别器之间的纳什均衡。在这样的理想状况下，鉴别器无法有效鉴别图像是否来自生成器或来自真实数据，只能随机猜测结果。GAN 模型被广泛应用于缺失多模态学习中，其主要思路是利用 GAN 生成缺失样本并进行后续处理与分析。然而，现有基于 GAN 的缺失多模态学习方法主要面向集中式数据处理场景，即所有数据均由同一实体持有。如何将 GAN 应用于联邦缺失多模态聚类，尚缺乏相关工作。

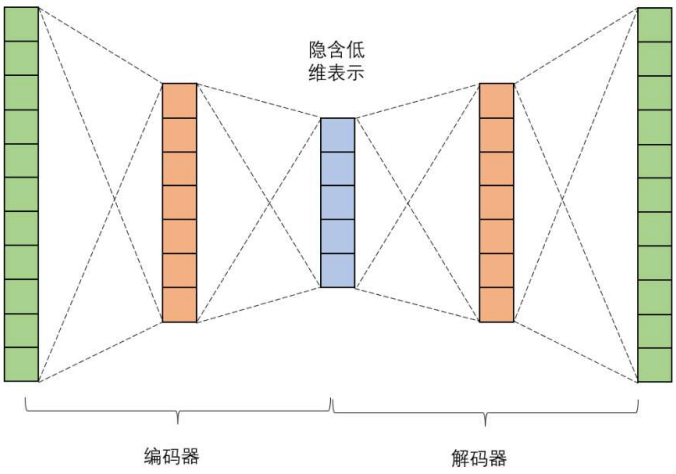


图 1.2　自编码器简略模型图

1.3 域映射

域映射试图从不配对的数据中构造和识别两种模态之间的映射，如果能够学习得到模态之间的映射关系，就可以用于多模态数据补全。因此一些学者提出许多域映射的方法。一些基于 GAN 的方法如下：Kim 等人[13]提出 DiscoGAN 能够使用嵌入对应于另一个域的自编码器模型发现跨域关系。生成器学习从一个域映射到另一个域，而另一个生成器将其映射回原始域。每个域都有一个鉴别器来辨别生成的图像是否来自本域，同时模型中还有一个重建损失确保映射的准确性。Zhu 等人[14]则以非常相似的方式训练一个称为循环对抗生成网络（Cycle Generative Adversarial Network，CycleGAN）的模型，其中 Cycle 表现为循环一致损失。pix2pix 方法[15]与 CycleGAN 类似，但是该方法是针对匹配数据进行训练，要求数据集提供两个域之间的准确匹配信息，因此数据集条件较为苛刻。由于域映射可以学习模态和模态之间的映射关系，从而基于某个模态数据生成缺失模态数据，因此在本文的模型中也有相应的应用。

1.4 自编码器及多模态自编码器

1.4.1 自编码器

自编码器（Autoencoder，AE）是一种神经网络模型，由编码器和解码器组成。编码器将输入数据压缩成低维表示即映射到特征空间，而解码器将该表示还原为原始数据。自编码器通过最小化公式(1.7)中所示的重构误差来学习数据的特征，并可用于数据压缩、特征提取和生成新样本。它能够自动学习数据的表示，无需标注数据，被广泛应用于数据降维、去噪等场景中，AE 的基本模型如图 1.2 所示。

$$min_{f,g}\left\|X-g\left(f(X)\right)\right\|^2 \tag{1.7}$$

自编码器也有许多拓展应用，如 Vincent 等人[16]提出的去噪自编码器(Denoising autoencoder，DAE) 就可以通过接收叠加有噪声的原始数据，来训练预测精确原始数据的模型，从而达到去除数据中噪声的作用。因此它的损失函数为：

$$min_{f,g}\left\|X-g\left(f(\tilde{X})\right)\right\|^2 \tag{1.8}$$

其中，$\tilde{X}$为X的某个加噪样本。

1.4.2 多模态自编码器

当将自编码器应用到多模态领域时，J.Ngiam 等人[17]提出多模态自编码器，一种双模态自编码器的模型图如图 1.3 所示。由于模态与模态之间是有依赖关系的，因此可以

从成对的模态之间潜在的共享特征中捕获模态之间的相关性，从而根据某一模态生成对应缺失模态的数据。但是由于该模型是基于匹配数据进行训练的，而在实际应用中，可能存在配对数据不足现象，这极大的限制了其实际应用。

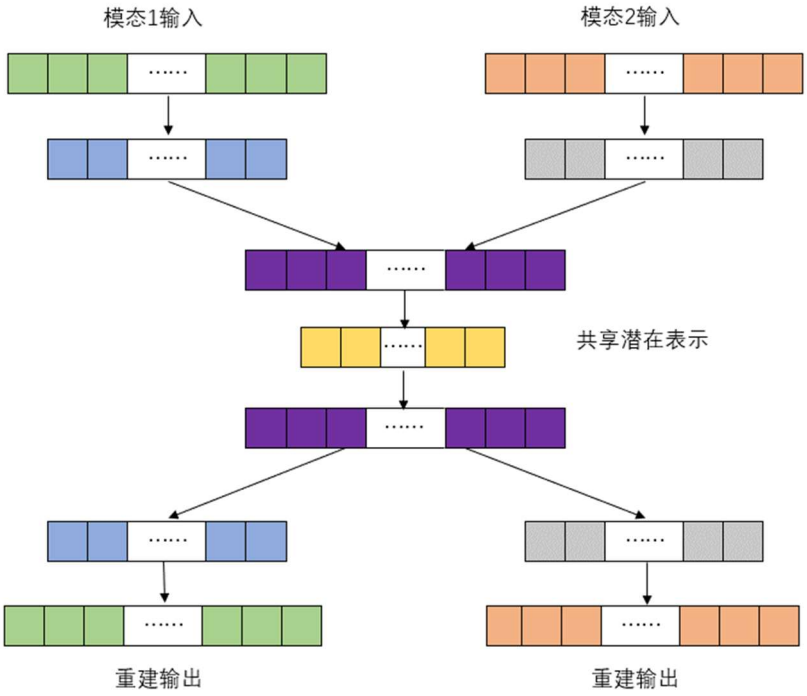


图 1.3　双模态自编码器简略模型图

1.5　联邦学习

联邦学习（Federated Learning，FL）[18]是一个机器学习框架，能有效帮助多个机构在满足用户隐私保护、数据安全和政府法规的要求下，进行数据使用和机器学习建模。联邦学习作为分布式的机器学习范式,可以有效解决数据孤岛问题,让参与方在不共享数据的基础上联合建模，能从技术上打破数据孤岛，实现数据协作。

针对数据集在客户端上的分布情况，可用将联邦学习分为三类：横向联邦学习（Horizontal Federated Learning，HFL）、纵向联邦学习（Vertical Federated Learning，VFL）以及联邦迁移学习（Federated Transfer Learning，FTL），如图 1.4 所示。其中，横向联邦学习是针对数据集之间特征重叠较多，而样本重叠较少的情况比如，而纵向联邦学习则是数据集共享样本空间，而特征彼此分离的场景，迁移学习面向数据集中特种重叠和样本重叠都较少的情况

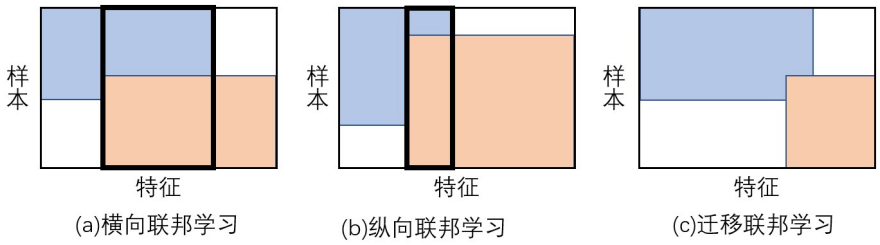


图 1.4 联邦学习分类

联邦学习的整个过程大致分成两个部分：模型训练和模型推理。在训练过程中，多方通过交换模型的信息来更新彼此的最新模型参数，通信过程可以采用加密的方式来保证模型信息的安全性。具体如下：假设在联邦场景中，有三个客户端 A、B、C，而中央平台则是一台聚合服务器，首先，中央服务器先将初始模型参数分发给 A、B、C，三个客户端分别使用本地的数据进行模型训练，然后将本地训练后的模型权重发送回中央服务器，之后中央服务器对这些模型权重进行更新聚合，并且将聚合的结果再次分发给各个参与方，重复整个过程。相比于集中训练需要暴露本地数据，联邦学习并不需要进行数据的集中处理，却能利用全局数据训练模型，在保护隐私的情况下也保证了最终训练的模型具有良好的可用性。

常见的联邦学习算法有联邦平均算法（Federated Average，FedAVG）[18]改进版本 FedProx （Federated Proximal）[19]等。在 FedAVG 算法中，服务器在每一个全局周期内，随机部分客户端上进行随机梯度下降（Stochastic Gradient Descent，SGD)，然后回收相应的梯度值进行在客户端进行全局的平均值的求值以作为全局模型的参数，并将更新后的权重重新发送回所有客户端。而在 FedProx 则致力于解决联邦学习中的非独立同分布（non-independent and identically distributed，non-IID）数据问题中，它在聚合客户端返回的梯度信息时向其中引入一个近似项，从而增大与全局模型近似的本地模型的权重，同时也能够协调考虑全局模型的泛化性。

1.6 本章小结

本章主要介绍了 GAN、域映射、自编码器及其扩展的多模态自编码器和联邦学习的概念和拓展知识，并对现有相关工作在处理缺失多模态数据时的优势与不足进行了简单的分析，该章内容为后续章节中 Fed-VIGAN 模型的提出奠定基本的理论基础与技术手段。

2 模型框架

在本章中，将会进行整体模型架构。在多源多模态数据缺失问题背景下，本文将基于前置知识提出扩展到联邦场景下的基于 VIGAN[21]的缺失多模态数据补全模型，即面向缺失数据补全的联邦生成对抗网络（Federated Generative Adversarial Networks for missing View Imputing， Fed-VIGAN）。本章首先在 2.1 节给出问题的形式化定义以及模型的整体框架和符号说明；接着，在 2.2 节和 2.3 节中描述了各客户端本地 VIGAN 模型和基于联邦学习的全局模型 Fed-VIGAN，最后在 2.4 节中进行本章小结。

2.1 问题描述与符号说明

本文提出的 Fed-VIGAN 模型致力于基于以上缺失多模态数据集分布在多源客户端的情况完成缺失多模态数据补全，然后完成多模态聚类。该模型将 VIGAN 部署在每一个客户端上，作为本地模型，该本地模型基于本地的缺失数据集进行模型训练。而 Fed-VIGAN 全局模型则利用 FedAVG 算法在称之为联合训练的第二步训练过重中聚合所有客户端的本地模型。Fed-VIGAN 适用于缺失多模态数据的场景。具体模型将在本节后文中详细介绍。

为了方便符号说明，本文假设所选用的数据集只有两个模态，因此实际的缺失问题可以由引言中的图 1 简化成的图 2.1 表示：

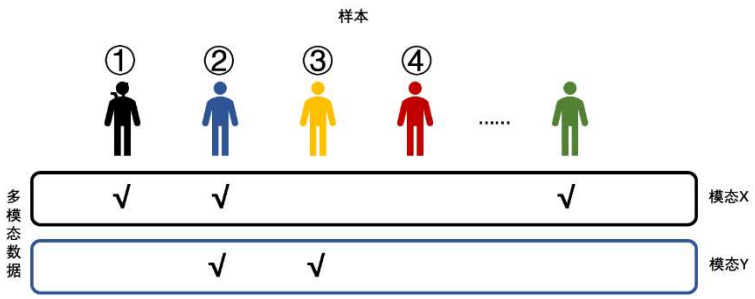


图 2.1 双模态数据缺失问题图例

从图中可以看到，样本①、样本③均存在数据缺失，而样本②则不存在数据缺失。于是可以定义类似样本②的模态 X 和模态 Y 上的一组数据为匹配数据（paired data），用(x, y)表示；而样本①的模态 X 数据和样本③的模态 Y 数据为一组不匹配数据（unpaired data）。

根据以上匹配数据和不匹配数据的定义，最终可以将给定的缺失数据集分为三个组成部分：所有匹配数据集合$\{(x_i, y_i)\}_1^N$，缺失对应模态Y数据的模态X数据集$\{x_i\}_1^{M_x}$以及缺失对应模态X数据的模态Y数据集$\{y_i\}_1^{M_y}$。基于以上问题定义，本文所提的 Fed-VIGAN 模型的整体目标是利用联邦学习机制训练 VIGAN 模型以补全各个客户端的缺失模态数据，并为后续学习任务（如聚类）提供支持。具体的，Fed-VIGAN 可分为两部分：本地模型与全局模型，其整体工作流程叙述如下：各个模型利用本地的缺失多模态数据训练本地模型，并利用 FedAvg 算法在服务器端进行参数融合得到全局模型。为了更好的介绍本文工作，本文在表 2.1 总结了与本文相关的符号并进行了说明：

表 2.1　模型符号说明

符号	说明
$G_1: X \to Y$	从模态X到模态Y之间的映射关系，或理解成基于模态X数据生成对应缺失模态Y数据的生成器
$G_2: Y \to X$	从模态Y到模态X之间的映射关系，或理解成基于模态Y数据生成对应缺失模态X的生成器
D_1	生成器G_1对应的鉴别器。
D_2	生成器G_2对应的鉴别器。
$A: X \times Y \to X \times Y$	多模态自编码器的单向生成
$E_{x \sim p_{data}(x)} f(x) = \frac{1}{M} \sum_{i=1}^{M} f(x_i)$	在相应分布上的算术平均
$P_X(X, Y) = X$	表示选定一组数据(x, y)中第一个模态X上的数据x。
$P_Y(X, Y) = Y$	表示选定一组数据(x, y)中第二个模态Y上的数据y。

2.2　本地模型

在本节中，将对本地模型 VIGAN 的网络结构训练过程的进行详细说明。整体来看，该网络由一个多模态自编码器（Multimodal AutoEncoder，Multimodal AE）和一个 CycleGAN 组成，其主要目的为利用本地缺失多模态数据训练一个本地生成模型，而不同客户端训练的本地模型将利用 2.3 节中的方法融合为全局模型。

该本地模型的训练过程可以分成三个步骤：第一步，基于匹配数据训练多模态自编码器用于学习模态间的潜在共享特征并且可以在最终的模型中演变为多模态去噪自编码器用于去噪处理；第二步，基于不匹配数据初步预训练 CycleGAN，通过学习模态间的映射关系用于缺失模态数据的生成；第三步，基于匹配数据同时对多模态自编码器和 CycleGAN 进行联合训练，由于此时多模态自编码器的输入是 CycleGAN 生成器的输出，

可以被认为是原始数据的加噪样本，因此最终多模态自编码器演变为多模态去噪自编码器进行工作。因此最终的模型图如图 2.2 所示，其中 CycleGAN 和多模态自编码器这两个单独的模块在最终的本地整体模型中有输入输出的耦合关系。

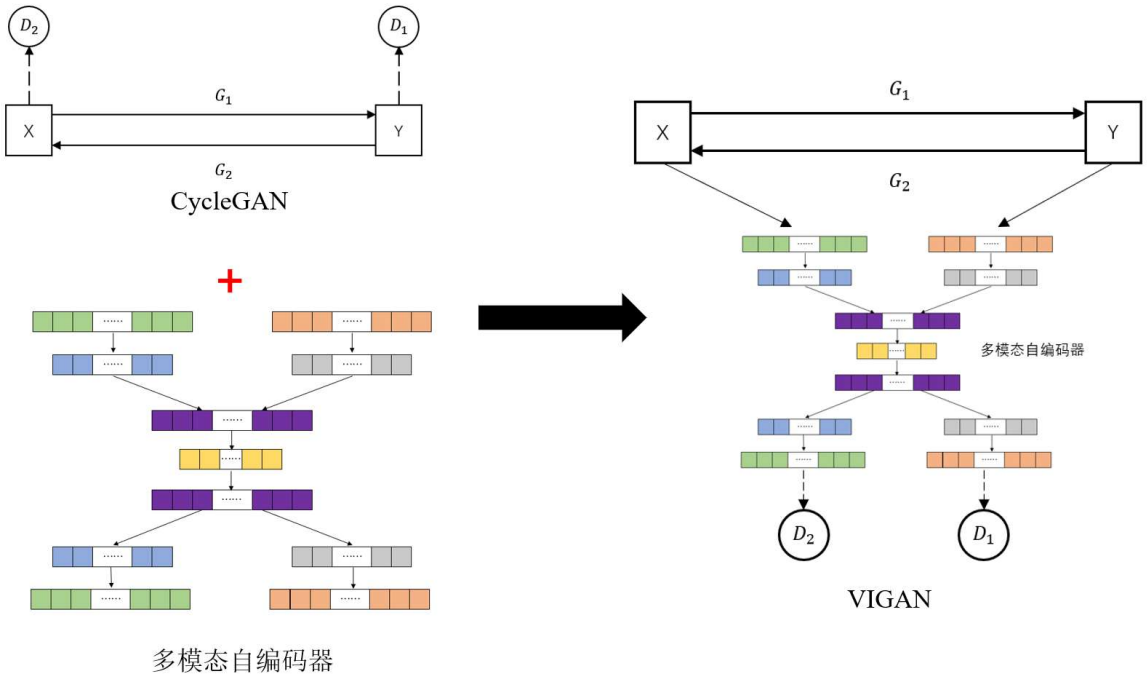


图 2.2　VIGAN 本地整体模型

2.3　全局模型

以上介绍了每个客户端需要维护的本地模型，但是当客户端本地数据量过小时或者需要利用全局信息来进行本地数据补全时，就可以考虑引入前文提出的联邦学习，利用 FedAVG 构建基于联邦学习的多源多模态数据补全模型 Fed-VIGAN，模型图如图 2.3 所示。

首先，在确定完本地客户端以及中央服务器维护的网络结构以及参数之后，选用 FedAVG 作为模型聚合的算法。在每一个全局周期结束后，中央接收本地客户端经过训练后返回的梯度信息并作参数平均，具体如下：

$$\theta_G^{Global}, \theta_D^{Global}, \theta_{AE}^{Global} = \frac{1}{N}\sum_{i=1}^{N}\theta_G^i, \frac{1}{N}\sum_{i=1}^{N}\theta_D^i, \frac{1}{N}\sum_{i=1}^{N}\theta_{AE}^i \tag{2.1}$$

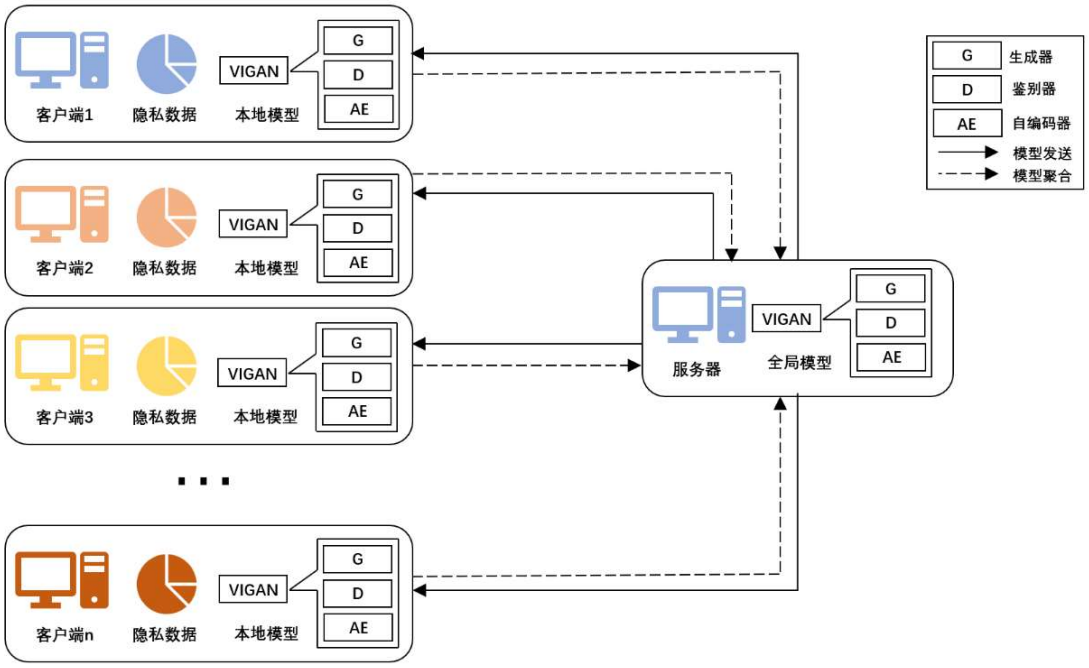


图 2.3　Fed-VIGAN 全局模型图

接着，以本地参数平均作为当前周期的全局参数，更新全局模型。最终将 Fed-VIGAN 用于模态生成构建完整数据集。

2.4　本章小结

在本章中，首先对本文拟解决的联邦缺失多模态问题进行了定义，基于此给出了 Fed-VIGAN 模型的整体框架设计，并详细介绍了本地客户端维护的 VIGAN 模型，包含其中的每个模块的组成、详细作用以及损失函数的构成和计算；然后，通过引入联邦学习框架将其推广至联邦学习场景，描述了其在联邦学习场景下的整体模型以及工作方式。

3 多源多模态缺失数据补全算法

在本章中，将会在 3.1 节和 3.2 节中对 Fed-VIGAN 的本地三阶段训练算法以及 Fed-VIGAN 的两阶段训练损失函数、算法以及训练过程进行详细介绍，并在 3.3 节中进行本章小节。

3.1 本地 VIGAN 算法

在本节中，将会对 VIGAN 的三阶段训练模型进行介绍，第一阶段为多模态自编码器的训练，第二阶段为 CycleGAN 的训练，第三阶段为融合两者的整体 VIGAN 训练。分别在对应的小节中会介绍相应的模块功能以及损失函数。

3.1.1 多模态自编码器

从图 1.3 中可以看出，多模态自编码器有多个输入和多个输出。它首先分别将每一个模态的高维度的模态输入转为低维表示，从而捕获模态的关键特征。接着将这些模态的关键特征嵌入到共享表示中，从而捕获模态间的相关性。最后再重新对每个模态数据进行重建。最终的目标是能够利用匹配数据中模态间的相关性进行数据重建。

多模态自编码器的损失函数在预训练中则体现为重建损失：

$$Loss_{AE}(A) = E_{(x,y)\sim p_{data}((x,y))}\|A(x,y) - (x,y)\|_2^2 \tag{3.1}$$

而当在第三步训练时，由于考虑到$G_1: X \to Y$和$G_2: Y \to X$分别是从模态X到模态Y之间的映射关系和从模态Y到模态X之间的映射关系，因此对于一组匹配数据(x,y)，则可以将G_1的输出$G_1(x)$视为对应y的近似版本，相应的，可以将G_2的输出$G_2(y)$视为对应x的近似版本。因此可以得到该匹配数据(x,y)的两个加噪样本，分别为$\big(x, G_1(x)\big)$和$(G_2(y), y)$，最终这两个加噪样本重建(x,y)。综合式(1.8)中的去噪自编码器的损失函数，可以提出在第三阶段多模态自编码器的损失函数，如式(3.2)所示：

$$\begin{aligned} &Loss_{AE}(A, G_1, G_2) = \\ &\quad E_{(x,y)\sim p_{data}((x,y))}\big\|A\big(x, G_1(x)\big) - (x,y)\big\|_2^2 \\ &+E_{(x,y)\sim p_{data}((x,y))}\|A(G_2(y), y) - (x,y)\|_2^2 \end{aligned} \tag{3.2}$$

3.1.2 CycleGAN

第二步是基于不匹配数据预训练一个 CycleGAN，训练后的 CycleGAN 用于缺失模态数据的生成，而该 CycleGAN 会在第三步中进行更加精确的训练，因此在本小节中将会介绍 CycleGAN 在第二步中训练时的整体损失函数。

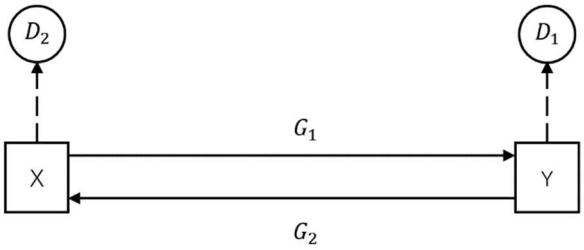


图 3.1　CycleGAN 模型图

CycleGAN 是一种无监督的图像转换模型，旨在不强制训练数据为匹配数据时也可以进行训练从而将一个域中的图像转换到另一个域中的图像。因此在本步中，可以利用由$\{x_i\}_1^{M_x}$和$\{y_i\}_1^{M_y}$构成的不匹配数据集来训练 CycleGAN。它通过两个生成器G_1和G_2以及对应的鉴别器D_1和D_2来实现这一目标，模型如图 3.1 所示。生成器G_1旨在将域X中的图像转换为域Y中的图像，而生成器G_2则相反，试图将域Y中的图像转换为域X中的图像。它们之间的转换是通过最小化两个关键的损失函数来实现的：对抗性损失和循环一致性损失。

对抗性损失通过训练鉴别器D_1和D_2来促使生成器产生更逼真的转换图像。鉴别器D_1旨在区分通过生成器G_1生成的域Y图像与真实的域Y图像，而鉴别器D_2则用于区分生成器G_2生成的域X图像与真实的域X图像。同时两个生成器的目标是欺骗鉴别器，使其无法区分生成的图像和真实的图像。因此对抗性损失定义为公式(3.3)：

$$\begin{aligned} Loss_{GAN} = Loss_{GA} \quad + Loss_{GAN2} = \\ E_{y\sim P_{data}(y)}\, log\big(D_1(y)\big) + E_{x\sim P_{data}(x)}\, log\left(1 - D_1\big(G_1(x)\big)\right) \\ + E_{x\sim P_{data}(x)}\, log\big(D_2(x)\big) + E_{y\sim P_{data}(y)}\, log\left(1 - D_2\big(G_2(y)\big)\right) \end{aligned} \tag{3.3}$$

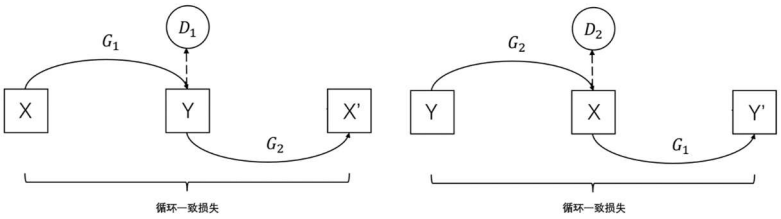


图 3.2　循环一致损失图示

如果只直接采用如上的对抗性损失，会导致如下问题：对于某一输入数据x，可能会生成域Y上任意一个关系不大的数据，但是模型在实际训练的时候是希望能够生成对应域内有隐含关系的数据，因此在以上损失函数的基础上，在考虑引入循环一致损失（Cycle Consistency Loss），该损失可以用图 3.2 描述。

基于该模型图以及定义，可以得到循环一致损失为：

$$Loss_{CCL}(G_1, G_2) = E_{x\sim P_{data}(x)} \left\| \left(G_2\left(G_1(x)\right)\right) - x \right\|_1 + E_{y\sim P_{data}(y)} \left\| \left(G_1\left(G_2(y)\right)\right) - y \right\|_1 \quad (3.4)$$

最终第二步的损失函数为对抗性损失和循环一致损失之和：

$$Loss_{CYC} = Loss_{GAN} + Loss_{CCL} \quad (3.5)$$

而在训练时，可以通过参数λ_1来改变对抗性损失和循环一致损失的权重，此时，整体损失函数为：

$$Loss_{CYC} = \lambda_1 Loss_{GAN} + Loss_{CCL} \quad (3.6)$$

CycleGAN 在第三步中继续被训练，但是损失函数有所变化，将在下一节中详细介绍。

3.1.3 VIGAN 算法

第三步是整合第一步和第二步中预训练得到的多模态自编码器和 CycleGAN 形成 VIGAN，并进行联合训练。整合后的模型结构如图 3.3 所示。

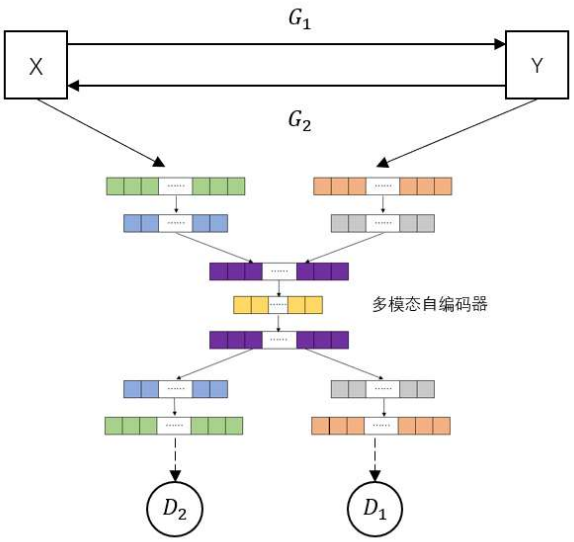


图 3.3 VIGAN 模型图

在该 VIGAN 模型中，以 CycleGAN 的生成器输出作为多模态自编码器的输入，具体表现为对于一个模态X中的一个数据x，通过生成器G_1生成对应另一个模态的假数据$G_1(x)$，因此$(x, G_1(x))$可以看作真实多模态数据(x, y)的预测值，但是其中不可避免含有一些误差，因此可以将其视为(x, y)的一个加噪样本，同理可得(x, y)的另一个加噪样本为$(G_2(y), y)$。因此此时多模态自编码器是为了基于加噪样本重建原始数据，故可以引入去噪自编码器的概念将其演变为多模态去噪自编码器，因此对应的损失函数如公式(2.2)所示。

而在其中，CycleGAN 的损失函数也有所变化，但是仍然由对抗性函数和循环一致损失两部分组成。由于本文假定的前提为数据只有两个模态，但是模型在一次缺失模态生成的过程中，只会产生一个新的模态数据，因此在确定损失函数的时候需要针对其中某一模态进行，也就是需要基于公式(3.4)和公式(3.5)进行修改。

其中确定对抗性损失，需要明确鉴别器实际上是需要对经过多模态自编码器去噪后的结果进行鉴别，而每个模态上都存在一个对抗性损失，因此对抗性损失由以下两部分组成：

$$Loss_{AG}^{Y}(A, G_1, D_1) = E_{y\sim P_{data}(y)} log\big(D_1(y)\big) + E_{x\sim P_{data}(x)} \log\Big(1 - D_1\Big(P_Y\big(A(x, G_1(x))\big)\Big)\Big) \quad (3.7)$$

$$Loss_{AG}^{X}(A, G_2, D_2) = E_{X\sim P_{data}(X)} \log\big(D_2(x)\big) + E_{y\sim P_{data}(y)} \log\Big(1 - D_2\Big(P_X\big(A(G_2(y), y)\big)\Big)\Big) \quad (3.8)$$

因此对抗性损失为公式(3.7)和公式(3.8)之和：

$$Loss_{AG}(A, G_1, G_2, D_1, D_2) = Loss_{AG}^{Y}(A, G_1, D_1) + Loss_{AG}^{X}(A, G_2, D_2) \quad (3.9)$$

接下来确定循环一致损失，由于循环一致损失主要是用于衡量经过两次生成器重建图像与原始图像的差异，因此公式(3.4)依然适用。

基于以上推导，合并公式(3.2)、公式(3.4)和公式(3.9)可以得到整体模型损失函数：

$$Loss(A, G_1, G_2, D_1, D_2) = Loss_{AE}(A, G_1, G_2) + Loss_{CCL}(G_1, G_2) + Loss_{AG}(A, G_1, G_2, D_1, D_2) \quad (3.10)$$

然后通过将三个超参数$\lambda_{AE}, \lambda_{CCL}, \lambda_{AG}$作为三个损失函数的权重在最终训练时达到更好的平衡效果，可以得到最终的整体模型损失函数为：

$$Loss(A, G_1, G_2, D_1, D_2) = \lambda_{AE} Loss_{AE}(A, G_1, G_2) + \lambda_{CCL} Loss_{CCL}(G_1, G_2) + \lambda_{AG} Loss_{AG}(A, G_1, G_2, D_1, D_2) \quad (3.11)$$

最终模型的解决转化为解决如下极大极小博弈问题：

$$min_G max_D Loss(A, G_1, G_2, D_1, D_2) \quad (3.12)$$

3.2 Fed-VIGAN 算法

根据 3.1 节中提到的 VIGAN 三步训练过程，分别为：第一步基于数据集中的匹配数据对多模态自编码器进行预训练，第二步为基于数据集中的不匹配数据对 CycleGAN 网络进行预训练，第三步为基于匹配数据对多模态自编码器和 CycleGAN 进行联合训练。而在 Fed-VIGAN 中，则将第一步和第二步作为模型的预训练作为一个部分，因此可以将整个本地 VIGAN 模型训练过程分成预训练和联合训练两个部分：第一步为客户端多模态自编码器以及 CycleGAN 的预训练，然后中央服务器对客户端预训练的结果进行聚合。第二步为基于 FedAVG 进行对多模态自编码器和 CycleGAN 的联邦训练。

第二步算法过程如下：

a) 在每个全局周期开始时，本地客户端基于 3.1 节中的本地算法进行本地模型训练。

b) 各个本地客户端将训练结果（梯度信息）发送会中央服务器。

c) 中央服务器基于公式(2.1)进行模型聚合，并将聚合之后的结果发送回本地客户端。

d) 重复步骤(a)-(c)。

根据以上 Fed-VIGAN 算法，即可构建基于全局数据构建面向多源缺失多模态数据补全的联邦生成对抗网络，用于完成本地多模态数据补全。

3.3 本章小结

在本章中，对集中式的多模态数据补全模型 VIGAN 和面向多源数据分布的联邦多模态数据补全模型 Fed-VIGAN 进行了详细的介绍，具体包括他们的损失函数、算法设计以及训练过程。在后文中将会对 Fed-VIGAN 的生成性能进行测试实验。

4 生成性能测试实验

在本章中，主要对所提模型的生成性能高进行了测试。实验主要以均方误差 MSE 和峰值信噪比 PSNR 为主要指标，分析集中式 VIGAN 和所提出的 Fed-VIGAN 的生成性能差异。接着还分析了 Fed-VIGAN 和本地少量数据集 VGAN 的生成性能差异，最后与扩展到联邦环境下的 pix2pix 也进行了对比实验。本章将在 4.1 中进行数据集、客户端数据集分配、评价指标的说明。在 4.2-4.4 中设计实验进行生成性能的分析，4.5 节设计对比实验进行生成性能分析。

4.1 实验设计

4.1.1 数据集

对于数据集，本文选用 MNIST（Modified National Institute of Standards and Technology）手写数字数据集，该数据集是一个广泛使用的计算机视觉数据集。该数据集由手写数字的灰度图像组成，涵盖了 0 到 9 的 10 个数字。其中图片样例如图 4.1：

图 4.1 MNIST 数据集实例

每个图像都是 28x28 像素的灰度图像（单通道），表示了一个手写数字。图像中的像素值介于 0 到 255 之间（训练时将会对其进行归一化，是其值介于 0 到 1 之间），其中 0 表示黑色，255 表示白色。每个图像都有一个相应的标签，表示图像中手写数字的真实值。标签是 0 到 9 之间的整数，对应于图像中所示的手写数字。MNIST 数据集被分为两个部分，即训练集和测试集。训练集包含 60000 个样本，用于模型的训练和参数优化。测试集包含 10000 个样本，用于评估模型的性能和泛化能力。同时每个数字类别

在整个数据集中都有相似数量的样本，这使得数据集是相对平衡的。这为后文给客户端分配数据集提供了方便。

将 MNIST 数据集中每张图片作为对应样本的一个模态，而将依照手写数字的边缘作为样本的另一模态，如图 4.2 所示。其中，通过将每张图片进行膨胀操作之后，与原图作差即可得到边缘图片。

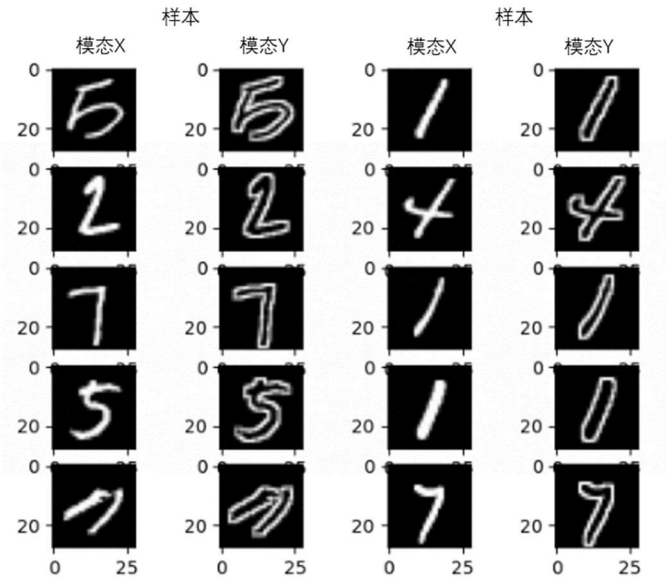


图 4.2　MNIST 数据集及对应边缘示例

在数据缺失问题中，缺失数据可以分为三种不同类型的产生缺失数据的机制[10]：

a) 数据完全随机缺失(data is missing at completely random，MCAR)，即数据的缺失完全随机发生，不依赖于任何变量

b) 数据随机缺失(data is missing at random, MAR)，即某个数据缺失的概率与其他一些变量的值有关

c) 数据不随机丢失(data is missing not at random，MNAR)，即数据丢失的概率与其值相关。

在本文中所讨论的多模态数据缺失问题是基于 MCAR 的，即多模态数据的丢失是完全随机的。因此可以构建基于以上多模态数据构建缺失多模态数据集。给定比例如表 4.1 所示：

表 4.1　训练集数据分配

	符号标识	数量
匹配数据	$\{(x_i, y_i)\}_1^N$	30000 对
模态 X 数据	$\{x_i\}_1^{M_X}$	15000 个
模态 Y 数据	$\{y_i\}_1^{M_Y}$	15000 个

4.1.2　客户端数据集分配

接着需要将上述数据集分配到多个客户端上。基于表 4.1 中的数据量，本文设定本地客户端个数 CLIENT_NUM=10，并且将以上三份子数据集均匀分配到本地如表 4.2 所示：

表 4.2　客户端训练数据分配

客户端编号	数据类别	符号标识	数量
客户端 1	匹配数据	$\{(x_i, y_i)\}_1^{N_1}$	3000 对
	模态 X 数据	$\{x_i\}_1^{M_{X1}}$	1500 个
	模态 Y 数据	$\{y_i\}_1^{M_{Y1}}$	1500 个
客户端 2	匹配数据	$\{(x_i, y_i)\}_1^{N_2}$	3000 对
	模态 X 数据	$\{x_i\}_1^{M_{X2}}$	1500 个
	模态 Y 数据	$\{y_i\}_1^{M_{Y2}}$	1500 个
…	…	…	…
客户端 10	匹配数据	$\{(x_i, y_i)\}_1^{N_{10}}$	3000 对
	模态 X 数据	$\{x_i\}_1^{M_{X10}}$	1500 个
	模态 Y 数据	$\{y_i\}_1^{M_{Y10}}$	1500 个

同时每个客户端满足独立同分布的条件，即每一个客户端中，每一个数字的占比是大致相等的，可以用公式(4.1)衡量。

$$\frac{Num(k,i)}{\sum_{s=1}^{9} Num\,(s,i)} \approx \frac{Num(k,j)}{\sum_{s=1}^{9} Num\,(s,j)}, 0 \le k \le 9, 1 \le i,j \le \text{CLIENT\_NUM} \tag{4.1}$$

其中$Num(k,i)$表示在客户端i上的标签为k的样本的个数

4.1.3　评价指标

本文设计的实验是基于缺失数据集，在联邦环境下进行数据补全，因此需要借助一些指标来衡量数据补全的性能。由于选用的 MNIST 手写数字数据集是图片数据集，因此可以采用一些图片质量的评价指标，如均方误差、峰值信噪比和结构相似性等。

均方误差（Mean Squared Error，MSE）是一种常用的图像评价指标，用于比较原始图像和重建图像之间的差异程度。它是计算两幅图像之间像素差异的平方和的平均值，即可以量化重建图像和原始图像之间的差异，并提供一个数值来表示它们之间的相似程度。如若试图通过 MSE 来衡量图片间的相似程度，则需要两张图片的大小相等。若两张图片I、I'的大小均为$m \times n$，则两者的 MSE 定义如公式(4.2)所示：

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j) - I'^{(i,j)}\right]^2 \tag{4.2}$$

MSE 的计算结果越小，表示重建图像与原始图像之间的差异越小，它们之间的相似程度越高。然而，MSE 也有一些缺点，例如对于亮度差异较大的图像可能会给出较大的误差值。因此，在某些情况下需要在结合其他的评价指标来全面评估图像的质量和相似度。

峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）是另一种衡量图像质量的指标，它通常用于比较原始图像和经过压缩或处理后的图像之间的差异。PSNR 的计算基于前文中介绍的 MSE。PSNR 的计算公式如下：

$$PSNR = 10log_{10}\left(\frac{L^2}{MSE}\right) \tag{4.3}$$

其中，L 表示像素值的最大可能取值（对于8位灰度图像来说，$L = 255$）。MSE 越小，PSNR 的值就越高，表示图像的重建质量越好。PSNR 的单位是分贝（dB），常用于度量图像的噪声水平和失真程度。较高的 PSNR 值意味着图像质量较好，而较低的 PSNR 值表示图像存在较大的失真或噪声。需要注意的是，PSNR 只能提供关于图像差异的定量信息，但不能反映人眼对这些差异的主观感知。

4.2 预训练—VIGAN 组成模块生成性能实验

在本节实验中，针对本地模型的预训练设计实验。将训练完的模型用测试集中的随机图像进行模态生成，根据图像生成效果分析本地 VIGAN 中多模态自编码器和 CycleGAN 的生成性能。

4.2.1 多模态自编码器模块

作为一个用于重建任务的自编码器，该多模态自编码器需要使得重建所得图像与原本图像基本一致，因此以公式(2.1)作为损失函数，在每个客户端本地基于匹配数据集训

练 20 轮。由于本地客户端是基于本地匹配数据进行的预训练，并没有和全局进行数据交互，因此全局衡量指标为客户端损失平均：

$$Loss_{AE}^{Global}(A) = \frac{1}{CLIENT\_NUM}\sum_{s=1}^{CLI\_NUM} Loss_{AE}^{S}(A) \tag{4.4}$$

其中$Loss_{AE}^{S}(A)$表示第s个客户端上多模态自编码器的损失。

每个客户端经过经过 20 轮的本地预训练，全局平均损失函数以及随机两个客户端的损失函数变化曲线如图 4.3 所示。

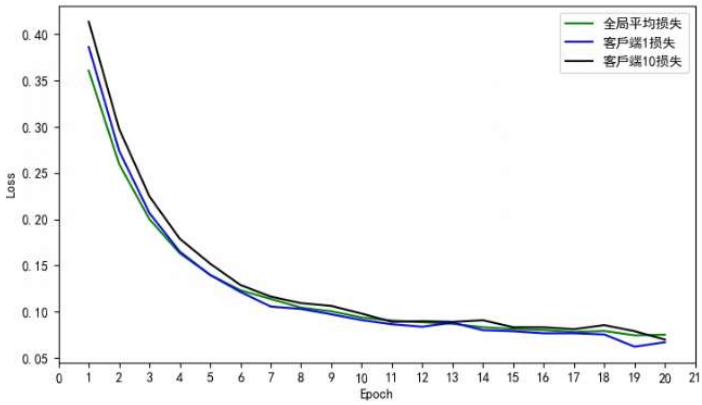


图 4.3　第一步多模态自编码器预训练损失函数变化曲线

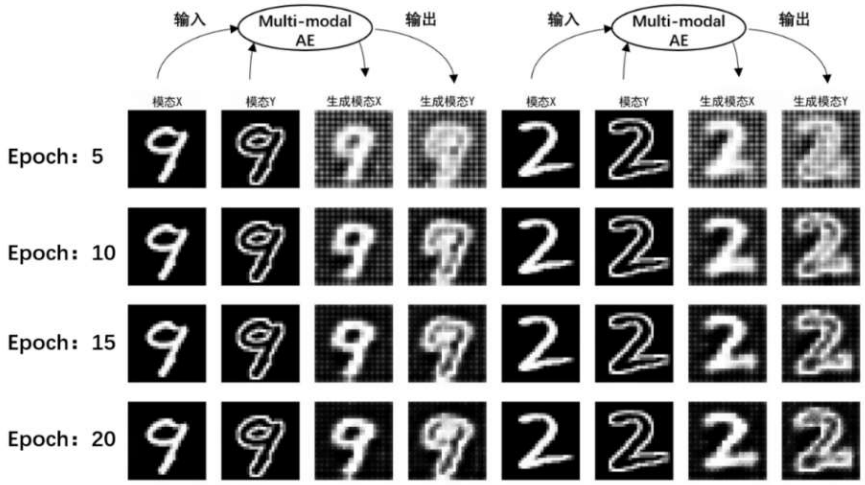


图 4.4　多模态自编码器训练 5，10，15，20 个周期后的重建效果

可以发现本地客户端的损失函数不断在减小的并最终趋平稳，也就是说明多模态自编码器的重建性能逐渐变好，这等价于该多模态自编码器能够捕获模态之间更深层的依

赖关系，并基于此重建图像。图 4.4 展示了依次训练 5，10，15，20 个周期之后，该多模态自编码器的基于随机样本的输入输出，可以发现输入输出得图片越来越相似，噪声越来越少，重建效果越来越好。图 4.5 为训练 20 个周期之后的重建效果。

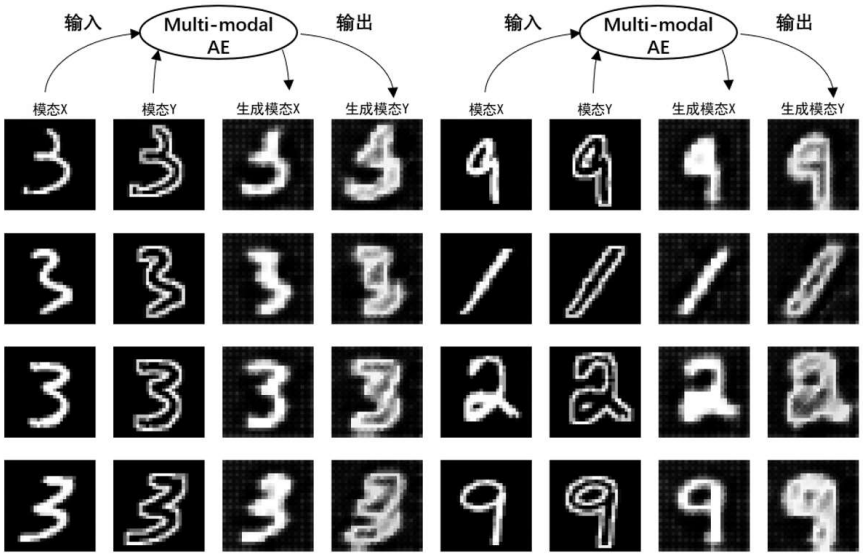


图 4.5　多模态自编码器训练 20 个周期后的重建效果

可以看到结果比较粗糙，但是由于多模态自编码器的重建效果比较一般，在经过 20 个周期的训练之后，生成的图像仍然比较模糊，质量较差。但是重建前后图像的隐层特征仍然没有丢失，重建后的两个模态仍然享有同样的特征，且与重建前的图像较为相似。同时考虑到该预训练的目的只是捕获模态之间的隐藏关系，为在联合训练中训练为去噪自编码器提供一定的参数基础，因此 20 轮训练出效果已经可以基本达成目的。

4.2.2　CycleGAN 模块

CycleGAN 预训练的目的是学习模态与模态之间的映射关系，它的特点是基于数据集中的非配对数据即可。经过训练之后的 CycleGAN 需要将基于一个模态的数据生成对应的缺失模态的数据，因此域映射的性能非常重要。与 4.2.1 类似，定于全局衡量指标为客户端的损失平均，定义如下：

$$Loss_{CYC}^{Global}(G_1, G_2) = \frac{1}{CLIENT\_NUM}\sum_{s=1}^{CLIENT\_NUM} Loss_{CYC}^{s}(G_1, G_2) \tag{4.5}$$

其中$Loss_{CYC}^{s}(G_1, G_2)$表示第s个客户端上 CycleGAN 的损失。

每个客户端经过 5 轮的本地预训练，全局平均损失函数以及随机两个客户端的损失函数变化曲线如图 4.6 所示：

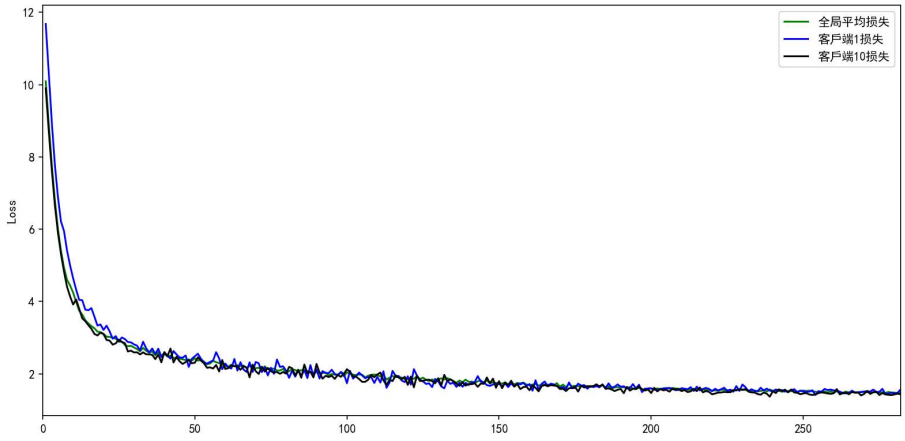


图 4.6　CycleGAN 训练损失函数变化曲线

可以发现本地客户端的损失函数不断减小并最终趋近平稳，也就是说明 CycleGAN 已经逐渐学习到了映射函数，因此可以用于缺失数据的初步补充。图 4.7 依次展示了训练了 1，2，3，4，5 个周期之后，CycleGAN 基于随机样本的某一模态生成对应缺失模态的图像。

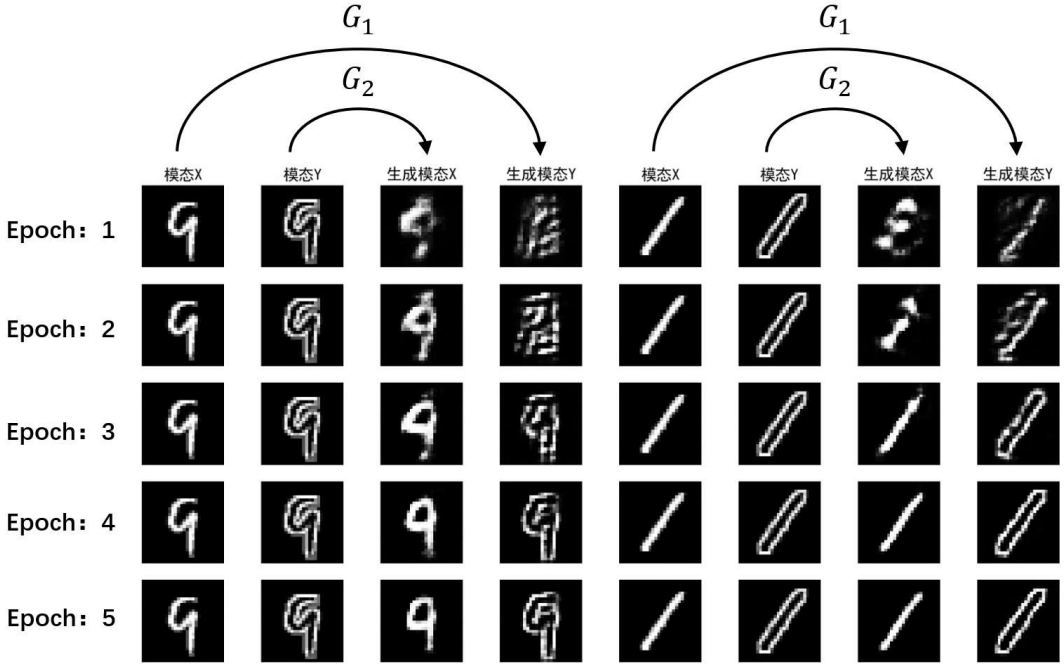


图 4.7　CycleGAN 训练 1，2，3，4，5 个周期后的生成效果

从图 4.7 中可以发现随着训练轮数增多，生成的模态图像与真实模态图像越来越相似，说明 CycleGAN 学习到的映射关系也越来越好。图 4.8 为训练 5 个轮次后生成数据的效果。

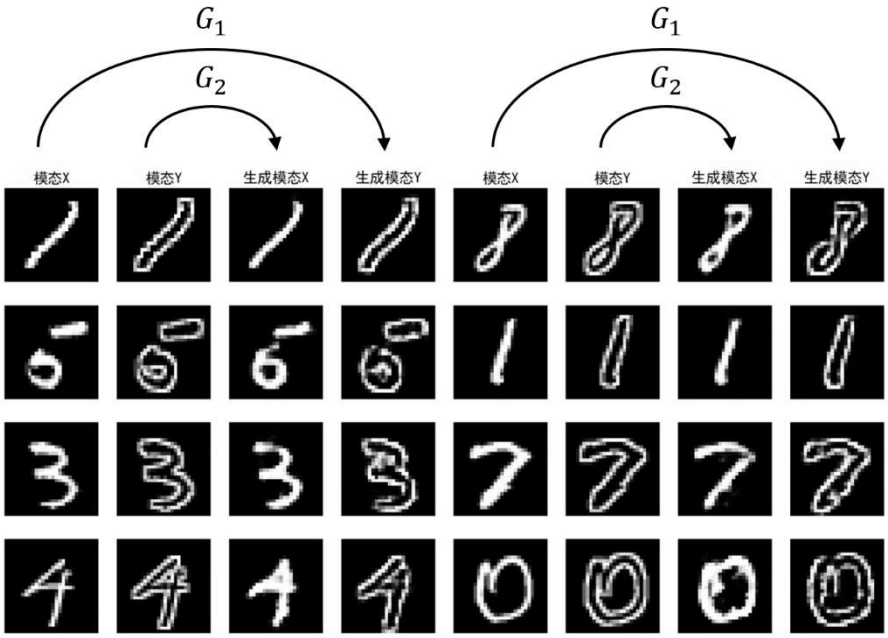


图 4.8　CycleGAN 训练 5 个周期后的生成效果

可以发现该 CycleGAN 只是经过了 5 轮的训练，整体的生成效果表现较好，如第二行第一组数据中，数字“6”的书写出现断笔，但是在生成另一模态的时候也能捕获到。同时相比于图 4.5 多模态自编码器的生成效果，CycleGAN 的生成的图像质量明显较好，噪声更少，能够更真实的生成另一缺失模态信息。但是生成的图像质量仍然有部分变形。

4.2.3　预训练结果分析

在预训练中，VIGAN 对多模态自编码器基于匹配数据集和 CycleGAN 基于非匹配数据集分别进行了训练，在最终的结果进行对比可以发现，CycleGAN 具有较好的模态图像生成的功能，因此在整体模型中将会作为主要的生成模块。

4.3　联合训练生成性能实验

在第二步联合训练中，在前一步本地预训练的参数基础上，引入了联邦学习框架。在每一个全局周期开始时，中央服务器会将模型参数发送给所有客户端，而客户端依据模型参数基于节 3.1 所示进行训练，然后将梯度信息返回给中央服务器，之后中央服务器进行参数聚合，重复以上过程。

因此在这一过程中，中央模型会在每个周期之后进行更新，但是由于全局模型没有训练而只是对本地客户端的模型进行平均，因此仍然用每个周期本地客户端的损失函数平均值来表示训练过程的变化。进行 40 轮的联合训练可以发现损失函数实在不断下降并最终趋近于稳定，也就是说明联合训练生成图像和真实图像越来越接近，因此可能可以用于缺失模态的补充。图 4.9 展示了分别训练 10，20，30，40 轮时的生成效果。

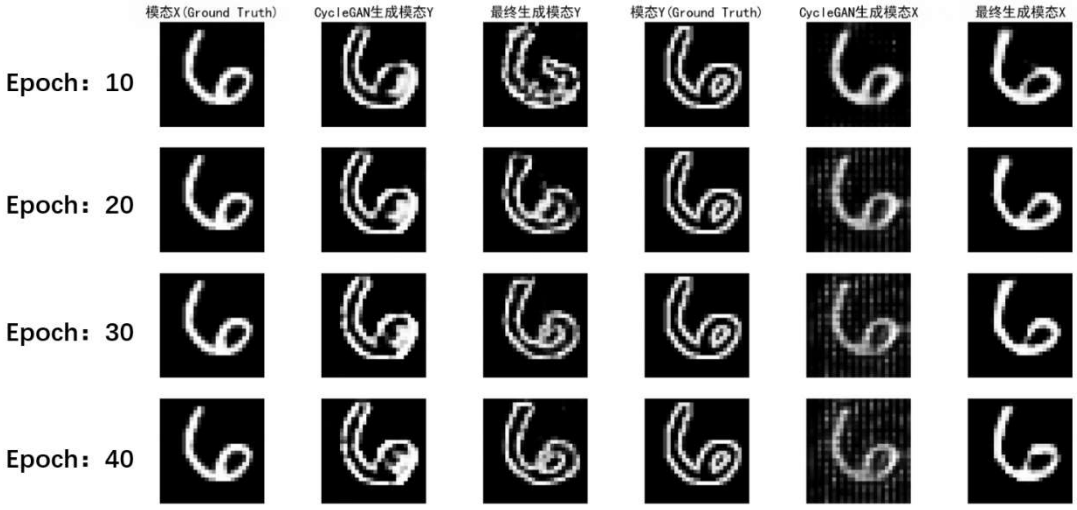


图 4.9　Fed-VIGAN 分别训练 10，20，30，40 个周期后的生成效果

随着训练轮数的增多，可以发现最终生成的模态与真实模态越来越相似，也就是说明整个模型的生成性越来越好。训练 40 轮后的一些生成效果如图 4.10 所示。

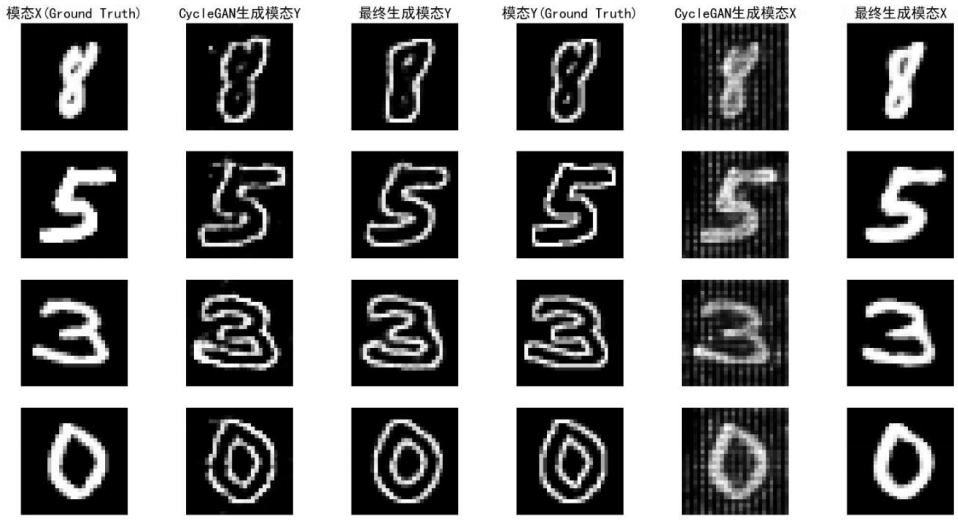


图 4.10　Fed-VIGAN 训练 40 个周期后的生成效果

在进行以下分析前再次明确一下模态生成的过程：假设现在存在X模态的图像，而对应的Y模态（边缘模态）有所缺失，首先将X模态的图像通过 CycleGAN 的$G_1: X \to Y$生成含有噪声的Y模态图像，然后将含有噪声的Y模态和原始的X模态图像通过多模态去噪自编码器即可生成最终的Y模态图像的近似。基于存在的Y模态图像生成缺失的X模态图像的过程类似。最终可以发现随着训练轮次的增加，模型中 CycleGAN 生成图像有更多的噪声，但是经过多模态自编码器进行降噪处理后生成的最终模态图像则于真实图像更为接近。

4.4 指标分析

以上的图片均形象化的展示了整体模型中各个模块的生成性能，但是没有直观数据进行数值判断，因此基于 4.1.3 中提到的衡量图片质量的量化评价指标对各个阶段性能进行实验测试与分析。

4.4.1 第一步预训练实验分析

多模态自编码器作用是捕获输入的多模态之间的潜在关系，从而重建出新的模态数据。根据 4.5 中的图片，可以发现该多模态自编码器的重建效果比较一般，生成的图像的噪声大、模糊、质量不好。基于 4.1.3 小节中的介绍，利用 MSE 和 PSNR 对其进行分析，结果如表 4.3 所示。

表 4.3　多模态自编码器训练时 MSE、PSNR 变化表

EPOCH	MSE	PSNR
1	21971.125	4.712
5	7345.722	9.471
10	5659.752	10.604
15	5229.051	10.946
20	4844.809	11.277

可以发现 MSE 随着训练轮数逐渐变大，而 PRNR 随着训练轮数逐渐减小，代表多模态自编码器重建图像的质量越来越好，最终训练 20 轮之后，MSE=4844.809，PSNR=11.277。

接下来分析 CyclgGAN 的生成性能。CycleGAN 的作用是基于某一模态数据生成另一缺失模态数据。根据图 18 的结果，可以发现生成图片的质量较好，噪声较少，因此可以作为整体模型的主要生成模块。基于 4.1.3 小节中的介绍，利用 MSE 和 PSNR 对其进行分析，结果如表 4.4 示：

表 4.4　CycleGAN 训练时 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	8184.782	9.001
2	4448.527	11.649
3	3260.162	12.998
4	2595.971	13.987
5	2026.374	15.064

同样的可以发现 MSE 随着训练轮数逐渐变大，而 PSNR 随着训练轮数逐渐减小，代表着 CycleGAN 生成模态数据的效果越来越好。最终训练 5 轮之后，MSE=2026.374，而 PSNR=15.064。

基于以上数据分析对预训练进行整体分析：多模态自编码器和 CycleGAN 作为整体模型的两个部分，在预训练时没有输入输出的关系，因此视为两个单独的模块进行分析。CycleGAN 作为域映射领域的一个基本模型，能够学习映射关系从而具有较好的生成图像的能力，而多模态自编码器可以在之后拓展为多模态去噪自编码器对图像进行去噪。

4.4.2　联合训练实验分析

在整体模型的第二部训练中，根据图 4.10 可以发现生成图片质量较高，与原图像差别较小。基于 4.1.4 章中的介绍，利用 MSE 和 PSNR 对其进行分析，结果如表 4.5：

表 4.5　Fed-VIGAN 训练时 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	4955.038	11.180
5	2458.345	14.224
10	2232.054	14.644
15	2168.372	15.075
20	2020.933	15.064
25	1881.179	15.387
30	1770.581	15.650
35	1748.304	15.705
40	1773.149	15.643

同样的可以发现 MSE 随着训练轮数逐渐变大，而 PSNR 随着训练轮数逐渐减小，代表着整体模型生成模态数据的效果越来越好。最终训练 40 轮之后，MSE=1773.149，而 PSNR=15.643。

4.4.3 整体分析

根据以上理论分析，可以发现在对多模态自编码器和 CycleGAN 在联邦环境下联合训练之后，就可以获得生成数据性能较好的整体模型，在后文将对模块的作用进行分析。

表 4.6 Fed-VIGAN 训练过程中，CycleGAN 模块的 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	2254.691	14.600
5	2576.486	14.021
10	3600.572	12.567
15	3894.330	12.226
20	4089.021	12.015
25	4018.345	12.090
30	4193.732	11.905
35	4592.809	11.510
40	5131.218	11.029

但是根据图 4.9 可以发现，在联合训练中，虽然本地 VIGAN 原始图像和生成图像的误差都越来越小，但是可以发现模块 CycleGAN 的输出图像中噪声越加明显，其 MSE 和 PSNR 变化如表 4.6 所示。

可以发现 CycleGAN 的 MSE 在不断增大，PSNR 在不断减小，也就是说明 CycleGAN 生成的图像中噪声越来越大，但是结合表 4.5 中最终的结果可以发现最终的 Fed-VIGAN 的可以使得补全的模态数据和真实数据更为相近。因此可以理解为在神经网络内部 CycleGAN 和多模态自编码器的均衡，从而使得整体网络的效果更好。

4.5 对比实验

Fed-VIGAN 能够在不进行客户端之间的样本数据交流的情况下，使得客户端基于全局数据补全本地缺失数据，其主要核心是基于联邦学习的模型训练。本节对数据集拥有方所拥有的数据集规模较小或者一个完整的数据集分别被多个数据拥有方持有分别设计实验进行分析。由于该问题存在两个方面，即数据集的物理性分离和单个数据集持有方拥有的数据量较小，可以分别将 Fed-VIGAN 与 VIGAN 基于全局数据集集中训练和 VIGAN 基于本地数据集本地训练进行对比实验。

4.5.1 集中式 VIGAN 训练对比

在该小节中，进行第一个对比实验的设置。将 MNIST 数据集集中在一个客户端，基于 VIGAN 的三步训练法进行训练，同时一张图片所有的超参数与 Fed-VIGAN 训练

保持一致。图 4.11 反映了在依次经过 10，20，30，40 次联合训练过程中，集中式 VIGAN 的生成性能。

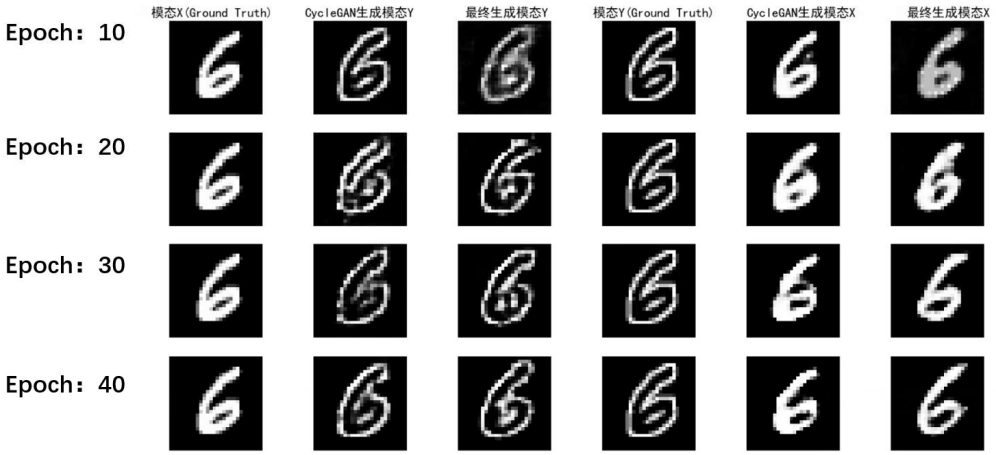


图 4.11　集中式 VIGAN 分别训练 10，20，30，40 个周期后的生成效果

以下表 4.7 基于 MSE 和 PNSR 对其性能进行分析：

表 4.7　集中式 VIGAN 训练时 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	3860.845	12.264
5	2187.176	14.732
10	1987.941	15.147
15	1872.982	15.405
20	1740.152	15.725
25	1663.809	15.920
30	1627.561	16.015
35	1563.130	16.191
40	1555.901	16.211

在完成 40 轮的集中式联合训练之后，MSE=1555.901，而 PSNR=16.211。与表 4.5 进行对比，可以发现在进行相同轮次的训练之后，集中式 VIGAN 相比于 Fed-VIGAN 具有更小的 MSE 和更大的 PSNR，这说明集中式 VIGAN 的性能较好于 Fed-VIGAN。

但是考虑到 Fed-VIGAN 是基于联邦学习的算法，主要目的是在保护隐私的前提下，基于全局数据集进行学习，因此生成性能稍弱于集中式 VIGAN 是可以接受的。Fed-VIGAN 可以在数据集分离的情况下利用联邦学习来获取全局的隐含信息，从而得到不次于集中式算法的性能。

4.5.2 本地 VIGAN 训练对比

在本小节中，进行第二个对比试验的设置。参照表 4.2 中本地数据集的分配，将 VIGAN 部署到某个客户端，基于该客户端的本地数据集进行本地 VIGAN 的训练，以分析 Fed-VIGAN 和数据量较少的 VIGAN 之间的模型性能差异。

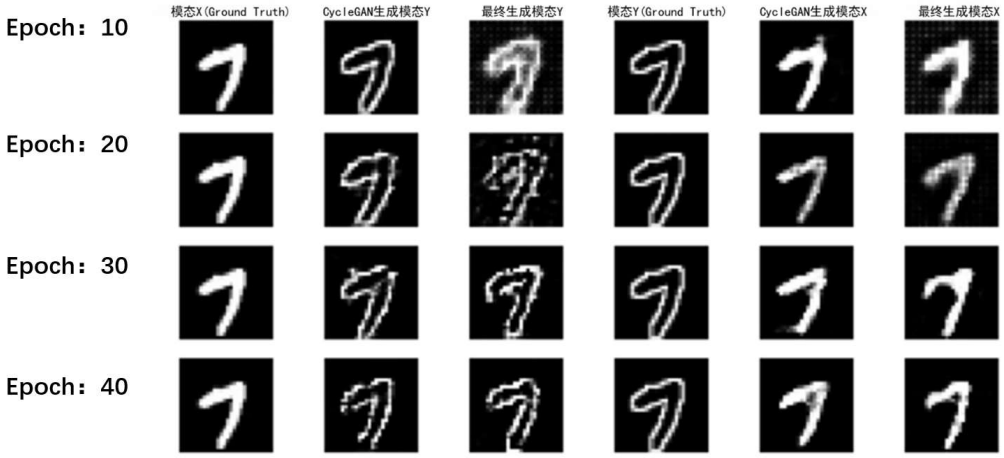


图 4.12 本地 VIGAN 分别训练 10，20，30，40 周期后的生成效果

图 4.12 反映了在依次经过 10，20，30，40 次联合训练过程中，本地 VIGAN 的生成性能。

表 4.8 基于 MSE 和 PSNR 对本地 VIGAN 性能进行分析：

表 4.8 本地少量数据集的 VIGAN 训练时 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	5111.133	11.045
5	4429.878	11.667
10	4306.743	11.789
15	3374.955	12.848
20	3109.872	13.203
25	2887.180	13.526
30	2913.811	13.486
35	2781.563	13.688
40	2676.680	13.854

在经过 40 轮次的训练之后，基于少量数据的本地 VIGAN 的 MSE=2676.680，PSNR=13.854。与表 4.5 对比可以发现，基于少量数据的本地 VIGAN 具有更大的 MSE

和更小的 PSNR，这说明 Fed-VIGAN 的生成性能较好于本地 VIGAN。这是因为 Fed-VIGAN 在训练的过程中结合了其余客户端的信息，因此生成性能更好。

4.5.3 pix2pix 对比

在对比实验中，选择选用一种比较传统的图像风格转换的方法 pix2pix，并将其应用到联邦场景下来进行对比实验。Pix2pix 和 VIGAN 相似，也是基于 GAN 的图像风格转换模型，从而扩展到图像数据补全方面。它的数据集要求是需要提供两个模态的匹配数据，从而学习模态间数据的映射关系。与 VIGAN 中的 CycleGAN 模块类似，pix2pix 的损失函数也分为对抗性损失和循环一致损失两个部分，整体损失函数如公式(4.6)、(4.7)、(3.8)所示：

$$Loss_{GAN}(G, D_Y, X, Y) = E_{y\sim p_{data}(y)}\|logD_Y(y)\| + E_{x\sim p_{data}(x)}\|1 - logD_Y(G(x))\| \quad (4.6)$$

$$Loss_{CYC}(G, F) = E_{x\sim p_{data}(x)}\|F(G(x)) - x\|_1 + E_{y\sim p_{data}(y)}\|G(F(x)) - y\|_1 \quad (4.7)$$

$$Loss_{pix2pix}(G, F, D_X, D_Y) = Loss_{GAN}(G, D_Y, X, Y) + Loss_{GAN}(G, D_X, X, Y) + Loss_{CYC}(G, F) \quad (4.8)$$

而与 CycleGAN 不同的是，它对鉴别器的输入是一组匹配的数据，即一组原始图像和生成器生成的图像或者一组真实的样本。在[15]中，pix2pix 被用于实景俯拍的街道线到地图风格街道线的转化、图片上色、物品边缘提取等领域，因此可以作为与本文工作对比比较。原本 pix2pix 是一个集中式的算法，类似将 VIGAN 扩展到联邦环境下的方法，本文将 pix2pix 也部署到联邦环境下。为了保持一致，在对比实验中也使用相同的客户端个数设置，相同的数据集 MNIST，相同的数据分布、相同的客户端本地数据集规模以及相同的联邦学习算法 FedAVG。每个客户端的本地数据表所示：

表 4.9 pix2pix 对比试验的客户端数据集分配

客户端编号	符号标识	数量
客户端 1	$\{(x_i, y_i)\}_1^{N_1}$	3000 对
客户端 2	$\{(x_i, y_i)\}_1^{N_2}$	3000 对
…	…	…
客户端 10	$\{(x_i, y_i)\}_1^{N_{10}}$	3000 对

基于同样的 FedAVG 对 pix2pix 进行 40 轮次的训练，pix2pix 输入输出指标分析随着轮次的变化情况如表 4.10 所示：

表 4.10　pix2pix 拓展到联邦环境下训练时 MSE、PSNR 变化

EPOCH	MSE	PSNR
1	7821.098	9.198
5	5634.761	10.622
10	4566.375	11.535
15	4223.936	11.874
20	3872.971	12.250
25	3508.650	12.679
30	2875.126	13.544
35	2475.994	14.193
40	2337.901	14.443

可以发现经过 40 轮的联邦训练之后，pix2pix 的 MSE=2337.901，PSNR=14.4431，而根据表 4.10 可得，在 40 轮的联邦训练之后，Fed-VIGAN 的 MSE=1773.149，而对应的 PSNR=15.643。该对比实验证明了 Fed-VIGAN 能够在更宽松的数据集条件下训练出性能更优的多模态数据补全模型。

4.6　本章小结

本章主要是对 Fed-VIGAN 的数据补全功能进行分析。主要基于 MNIST 数据集介绍了其基于图像的模态数据补全的性能。分别分析了 Fed-VIGAN 的本地模型组成、联邦训练过程，并且以均值方差 MSE 和峰值信噪比 PSNR 为评价指标对其性能进行分析。最终以传统的图像风格转换网络 pix2pix 作为对比模型，证明了 Fed-VIGAN 能够在更加宽松的数据集条件下依旧表现良好。

5 聚类实验

对缺失多模态数据进行联邦补全的目的是为后续应用提供完备模态的数据支持。考虑到多模态数据的海量性与复杂性导致的数据标注难度大、成本高等问题，本章主要针对聚类这一典型的无监督多模态学习方法进行研究。具体的，本文以缺失数据为基础，利用 Fed-VIGAN 对其进行数据补全，基于补全的数据实现多模态数据聚类，通过测试聚类效果进一步验证该模型是否能够为下游任务提供数据支撑。为更好的测试 Fed-VIGAN 模型的有效性，本章实验采用 K-means 这一基础聚类算法并分析其聚类性能。具体的，本章在 5.1 节中进行聚类实验设置的说明，并在其余部分展示了聚类性能的测试结果。

5.1 聚类设置

5.1.1 聚类评价指标

在聚类实验中，需要用一些聚类指标对聚类的性能效果进行直观的分析，包括聚类纯度、调整兰德系数、互信息和标准互信息，本小节分别给出了这些指标的定义与说明。

聚类纯度（Purity，P）定义为聚类正确的样本数除以总样本数，如公式(5.1)所示：

$$P = \frac{M}{N} \tag{5.1}$$

其中M、N分别表示聚类正确的样本数和总样本数。但是在实际的聚类实验中，无法明确知道每一个样本属于哪一类，而只能知道那些样本属于同一个类，将会导致M值无法直接求得。因此在实际计算过程中，只能通过计算最大值的方法间接得出。即判断聚类后每一个簇中，属于原有哪一类数据的样本最多，就可以将该簇视作这一类的聚类结果，因此该类中聚类正确的样本就是两者的交集。纯度的计算公式如下：

$$P(\Omega, C) = \frac{1}{N}\sum_{k} max_j \left|\omega_k \cap c_j\right| \tag{5.2}$$

其中N表示样本总数，ω_k表示聚类后的每一个簇中的样本的集合，$\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_K\}$表示聚类后簇的集合，$C = \{c_1, c_2, c_3, \dots, c_k\}$表示真实的分类情况，$c_j$表示第 j 类中的样本。最终的聚类纯度的值介于[0,1]，值越大则代表聚类的效果越好。可以发现由于该指标需要基于原有的分类情况计算，因此适用于有标签数据。

调整兰德系数（Adjusted Rand Index，ARI）与分类纯度有类似的地方。首先给定两个向量$X=[a_1,a_2,a_3,\dots a_s]$和$Y=[b_1,b_2,b_3,\dots b_s]$分别表示聚类后后簇的集合以及样本的真实分类情况，$s$表示簇的个数。因此可以获得联列表如图 5.1 所示：

X \ Y	Y_1	Y_2	…	Y_s	SUMS
X_1	n_{11}	n_{12}	…	n_{1s}	a_1
X_2	n_{21}	n_{22}	…	n_{2s}	a_2
…	…	…	…	…	…
X_s	n_{s1}	n_{s2}	…	n_{ss}	a_s
SUMS	b_1	b_2	…	b_s	

图 5.1　聚类结果和真实样本分类的联列表

其中n_{ij}表示X_i和Y_i的交集的个数。基于以上联列表可以得到 ARI 的计算公式（公式 5.3）。ARI 的值介于$[-1,1]$之间，值越大则代表聚类结果越好。

$$ARI=\frac{\sum_{i=1}^{s}\sum_{j=1}^{s}\binom{n_{ij}}{2}-\frac{\sum_{i=1}^{s}\binom{a_i}{2}\sum_{j=1}^{s}\binom{b_j}{2}}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_{i=1}^{s}\binom{a_i}{2}+\sum_{j=1}^{s}\binom{b_j}{2}\right]-\frac{\sum_{i=1}^{s}\binom{a_i}{2}\sum_{j=1}^{s}\binom{b_j}{2}}{\binom{n}{2}}} \tag{5.3}$$

互信息（Mutual information，MI）可以通常用于衡量聚类结果与数据集真实情况的相似度。同样基于图 5.1 中的联列表可以得到 MI 的定义：

$$MI(X,Y)=\sum_{i=1}^{S}\sum_{j=1}^{S}p_{ij}log\left(\frac{p_{ij}}{p_i\times q_j}\right) \tag{5.2}$$

且

$$p_i=\frac{a_i}{N},q_i=\frac{b_i}{N},p_{ij}=\frac{n_{ij}}{N} \tag{5.3}$$

其中a_i表示聚类结果中第i簇的样本数，b_j表示真实数据集中第j类的样本数，而$p_{ij} = \frac{|a_i \cap b_j|}{N} = \frac{n_{ij}}{N}$。互信息的值越大，则代表聚类结果与数据集真实情况越相似。

由于互信息的取值是越大越好的，即是没有上界的。因此为了更直接的显示聚类效果，引入标准互信息（Normalized Mutual Information，NMI）对其进行衡量。NMI 的定义为：

$$NMI(X,Y) = \frac{2MI(X,Y)}{H(X) + H(Y)} \tag{5.4}$$

其中$H(X), H(Y)$分别为X, Y的熵，定义为：

$$H(X) = -\sum_{i=1}^{s} p_i log p_i \, ; H(Y) = -\sum_{j=1}^{s} q_i log q_i \tag{5.5}$$

而p_i, q_j的定义与互信息中的定义相同。

NMI 的取值范围为[0,1]，值越大则代表聚类效果与真实数据集分布情况接近，说明聚类效果越好。

5.1.2 聚类实验设计

本小节中，将采用 K-means 聚类算法对图像数据进行聚类。主要针对以下四种情况分别进行聚类，判断聚类性能。

a) 真实数据集：MNIST 数据集和它对应的边缘数据集，在该设置下，数据的各个模态是完备的。

b) Fed-VIGAN 生成数据集：该数据集源于表 4.2 中的数据集分配，通过训练之后的 Fed-VIGAN 网络补全缺失数据从而得到的完整数据集。

c) 集中式 VIGAN 生成数据集：该数据集仍然源于表 4.1 中的数据集分配，但是通过集中训练之后的 VIGAN 网络补全缺失数据从而得到完整数据集。

d) 本地 VIGAN 生成数据集：该数据集基于表 4.2 中客户端 5 的数据集分配，数据量较少，为整体数据集的$\frac{1}{10}$，同时通过基于本地少量数据训练而成的本 VIGAN 网络补全缺失数据而得到的本地完整数据集。

对应的数据集每个样本都是包含他的手写数字图像以及边缘数字图像信息，聚类算法需要基于这两个信息进行聚类，并在 5.2 节中会根据 5.1.1 中介绍的聚类指标进行分析。

5.2 聚类性能

基于 5.1.2 中的联邦实验，本文使用 K-means 聚类算法对其进行聚类，聚类纯度 P 和调整兰德系数 AIR 指标：

表 5.1 四种聚类设置下聚类纯度和调整兰德系数结果

	真实数据集	Fed-VIGAN 生成数据集	集中式 VIGAN 生成数据集	本地 VIGAN 生成数据集
聚类纯度	0.496	*0.415*	**0.437**	0.277
调整兰德系数	0.299	*0.207*	**0.249**	0.078

由于互信息 MI 和标准互信息 NMI 可以衡量两个聚类结果之间的聚类效果的相似性，因此如果与真实数据集的聚类结果对比，则可以判断聚类结果与真实情况的相似程度。于是将聚类设置中 2、3、4 分别相于 1 求互信息和标准互信息，结果如表 5.2 所示：

表 5.2 聚类设置中 2、3、4 三种情况与情况 1 的互信息和标准互信息

	Fed-VIGAN 生成数据集合	集中式 VIGAN 生成数据集	本地 VIGAN 生成数据集
互信息	*1.041*	**1.209**	0.830
标准互信息	*0.329*	**0.383**	0.253

在以上的表格中，使用粗体来更加明显的表示其中聚类性能最好的，而用斜体表示我们提出的 Fed-VIGAN 算法的聚类性能，具体的分析将在后文进行。

5.3 聚类结果分析

根据表 5.1 中，可以发现普遍聚类纯度以及调整兰德系数都较小，这本质是因为 K-means 聚类算法的性能一般，因此本文进行相对性比较来分析最终的聚类结果。可以发现真实数据集的聚类结果的聚类纯度和调整兰德系数都是最高的，而集中式 VIGAN 生成的数据集在 K-means 下聚类结果的聚类纯度和调整兰德系数几乎与真是数据集的聚类结果相似，这说明集中式算法表现良好。而可以发现在 Fed-VIGAN 下对应聚类结果的聚类纯度和调整兰德系数与集中式 VIGAN 也非常接近，这说明了 Fed-VIGAN 在数据补全时能够有效学习到全局的数据特征进行补全，最终的聚类结果与真实数据集情况相似。而在表 5.2 中的互信息与标准互信息的数据也能证明以上结论。

5.4 本章小结

在本章中，通过聚类指标的分析来判断该 Fed-VIGAN 网络能否学习全局数据的信息，最终发现 Fed-VIGAN 有不弱于集中式 VIGAN 算法的性能，在聚类中表现良好。这证明 Fed-VIGAN 可以利用联邦学习框架保护隐私同时有效地学习多源数据的分布，并基于 VIGAN 的生成性高效地完成缺失多模态数据补全。

结　论

本文针对基于联邦学习的缺失多模态学习问题进行了研究，提出了面向多源缺失多模态数据补全的联邦生成对抗网络（Fed-VIGAN），通过利用 FedAvg 算法，将集中式的缺失多模态生成模型 VIGAN 拓展到联邦场景下，构建出一个两阶段的联邦缺失多模态生成模型。本文以两个模态的场景对模型进行了阐述，但 Fed-VIGAN 可以扩充至多模态领域，完成缺失多模态聚类。为验证模型的有效性，本文在 MNIST 手写数字数据集上进行实验，综合利用生成数据可视化以及对生成数据的一系列指标测试验证了 Fed-VIGAN 的生成效果；通过将 Fed-VIGAN 和集中式 VIGAN 的生成效果进行对比发现，其效果与集中式的 VIGAN 性能接近，进一步说明了 Fed-VIGAN 在限制数据交互的情况下学习多源数据集数据分布。最后，本文基于 K-means 聚类算法设计聚类实验可以证明 Fed-VIGAN 这一联邦缺失多模态能够为下游的无监督学习任务提供数据支持。

尽管本文实现了联邦学习场景下的缺失模态数据生成任务，但该方案仍存在局限性。首先，本文所提方案旨在扩充不同客户端的样本数量，没有充分考虑不同客户端所持有的多模态数据的异构性，其有效性一定程度上与不同客户端所持有的多模态数据分布的一致性相关；另一方面，在保护隐私方面，本文采用的联邦学习方法保护隐私的主要手段是避免数据的直接共享，虽然支持对模型参数进行加密，但对于反演攻击等隐私攻击方法的方法，没有充分考虑。在未来工作中，将进一步针对这些工作进行研究，以进一步提高模型的有效性与隐私保护能力。

参考文献

[1] Flanagan A, Oyomno W, Grigorievskiy A, et al. Federated multi-view matrix factorization for personalized recommendations[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases, European Conference, Ghent, Belgium, September 14–18, 2021:337-346.

[2] Huang M, Li H, Bai B, et al. A federated multi-view deep learning framework for privacy-preserving recommendations[J]. IEEE Transactions on Knowledge and Data Engineering, 2020,13:335-345.

[3] Che S, Kong Z, Peng H, et al. Federated multi-view learning for private medical data integration and analysis[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(4): 1-23.

[4] Zhao Y, Barnaghi P, Haddadi H. Multimodal federated learning on iot data[C]// Proceedings of the 2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI), F, 2022:43-54[C]. IEEE.

[5] Zhou X, Liu X, Lan G, et al. Federated conditional generative adversarial nets imputation method for air quality missing data[J]. Knowledge-Based Systems, 2021, 228: 107261.

[6] Yang B, Kang Y, Yuan Y, et al. ST-LBAGAN: Spatio-temporal learnable bidirectional attention generative adversarial networks for missing traffic data imputation[J]. Knowledge-Based Systems, 2021, 215: 106705.

[7] Zhang Y, Qu H, Chang Q, et al. Training federated gans with theoretical guarantees: A universal aggregation approach [J]. International Conference on Learning Representation, 2021, 11: 1103-1112.

[8] Xie G, Wang J, Huang Y, et al. Fedmed-gan: Federated domain translation on unsupervised cross-modality brain image synthesis[J]. Neurocomputing, 2023, 546: 126282.

[9] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy[C]// Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, Vienna, Austria, 2016:789-798.

[10] Rubin D B. Inference and missing data[J]. Biometrika, 1976, 63(3): 581-592.

[11] Lee D, Seung H S. Algorithms for non-negative matrix factorization[J] Advances in neural information processing systems, 2000, 13:545-556.

[12] Van Loan CFJSJonA. Generalizing the singular value decomposition[J]. SIAM Journal on numerical Analysis, 1976, 13(1):76-83.

[13]Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//Proceedings of the International Conference on Machine Learning, Sydney, Australia, 2017:1109-1121.

[14]Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision, Venice, Italy, 2017:2223-2232.

[15]Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017:1125-1134.

[16]Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 2008:1096-1103.

[17]Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning[C]//Proceedings of the 28th international conference on machine learning (ICML-11), Barcelona, Spain, 2011:689-696.

[18]McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the Artificial intelligence and statistics, Fort Lauderdale, Florida, 2017:1273-1282. PMLR.

[19]Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks [J]. Proceedings of Machine learning and systems, 2020, 2: 429-50.

[20]Corinzia L, Beuret A, Buhmann J M. Variational federated multi-task learning [J]. International Conference on Learning Representations, 2019, 9:678-686.

[21]Shang C, Palmer A, Sun J, et al. VIGAN: Missing view imputation with generative adversarial networks[C]//Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, Massachusetts, 2017:766-775.

修改记录

一、毕业论文题目修改

第一次修改记录：

原题目：基于多模态的联邦学习方法研究

修稿后题目：基于联邦学习的多模态聚类方法

二、毕业论文内容重要修改记录

第二次修改记录：

第 9 页 1.5，**修改前：**分为 1.5.1 和 1.5.2 两个小节

修改后：将两个小节合并

第三次修改记录：

第 11 页 2.1，**修改前：**对整体框架的过程介绍不够清晰

修改后：添加对整体框架的详细介绍

三、毕业论文外文翻译修改记录

第四次修改记录：改变摘要中的部分说法和措辞

四、毕业设计（论文）正式检测重复比：4.5%

记录人（签字）：武振威

指导教师（签字）：孙景昊

致　　谢

本次毕业论文是我人生第一次进行系统的科研，中间的磕磕绊绊不在少数。从十二月初的选题，再到一月份的正式开始毕业论文工作，到最后五月二十四日定稿，中间经历了也不少的挫折和失败，但是最终研究出一些成果。在校外指导老师冯伟老师和校内指导老师孙景昊老师的指导下，我逐渐对联邦学习、生成对抗网络这些概念有了基本的了解，对研究方向逐渐有比较清楚的认识。在科研工作中，代码编写的重要性是无言而语的，因此我在这一过程中，也将提高代码能力视为我自己的目标，在中途学习相关知识时，也注重模块的代码编写并进行尝试，自己的代码水平也在这过程中也略有提升。在整个毕业论文的过程中，碰到许多困难同时也发现了自己很多不足之处：比如不能明确自己的方向，在查找资料上的效率和准确性上有所欠缺、研究效率较低、动手能力较差等等。但是在整个过程中，在两位指导老师的悉心指导下，慢慢攻坚克难，逐步完成自己的工作。因此首先需要感谢的就是两位指导老师的悉心指导，他们通过此次毕业论文传授的经验方法将在日后的科研工作中继续发挥强大的指导作用。

接下来还要感谢我的大学同学们，我们在大二的时候一起从几个不同的专业转入计算机专业， 开始了我们一起竞争一起学习的三年。在这三年中，我们互相学习，互相讨论，在一次次的大作业和期末考试之前互相鼓励支持，最终都有了不错的结局和未来的发展方向。感谢大家在三年内对我的照顾和在毕业论文工作中的帮助。

接下来还要感谢在大学四年碰到的所有的老师们，为我大学所建立的人生观、世界观的建立都提供了宝贵的引领作用。也感谢计算机专业的所有老师，将我带入计算机专业学习的殿堂，你们的尊尊教诲我将铭记终生。还需要单独感谢王璇同学，在大学的最后两年我们相识相知，互相陪伴学习成长，以后也要一起一步一个脚印往前走。

最后还要感谢我的家人们，在我十余年的求学生涯中鼎力相助。特别是大学四年，我远离家独自在外求学，但是家人的关心一致围绕在我的身边，给予我最大的温暖和在陷入低谷时的助力。

漫漫求学路，需要的是坚持和努力作伴。初探科研的大门，体验到了一丝困难，迎难而上才是真正的选择。在未来的生活里，怀揣的男儿的梦想，继续努力。