分类号：TM39　　学校代码：10363
密　级：公开　　学　　号：2210330153

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目 虚假数据注入下智能电网攻击检测与容侵控制研究

论文作者 王瑞仁
校内导师 魏利胜（教授）
校外导师 郭春淼

专业学位类别 电子信息
研究方向（领域） 智能信息处理及应用

论文提交日期：2024 年 5 月 31 日

分类号：TM39　　单位代码：10363
密　级：公开　　学　　号：2210330153

题　目　虚假数据注入下智能电网攻击检测与容侵控制研究

题　目　Research on Attack Detection and Intrusion Tolerance Control of Smart Grid under False Data Injection

学　生　姓　名：王瑞仁
校　内　导　师：魏利胜教授
校　外　导　师：郭春淼
专业学位类别　：电子信息
研究方向（领域）：智能信息处理及应用

论文答辩日期：2024 年 5 月 12 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：王瑞仁

日期：　　2024　年　5　月　28　日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 √。

学位论文作者签名：王瑞仁　　指导教师签名：魏利胜

日期：　2024 年 5 月 28 日　　日期：　2024 年 5 月 28 日

虚假数据注入下智能电网攻击检测与容侵控制研究

摘　　要

智能电网是利用先进技术实现电力系统高效、安全、可靠、环保和经济运行的现代化电网系统。而虚假数据注入攻击(False Data Injection Attack，FDIA)是针对智能电网最具破坏性的恶意攻击类型之一，其通过设计攻击向量篡改量测数据，躲避能量管理系统(Energy Management System，EMS)中状态估计的不良数据检测器(Bad Data Detection，BDD)，引起调度中心对当前电网状态的误判，从而达到破坏电力系统安全稳定或获取不法收益的目的。因此，开展 FDIA 攻击检测与容侵控制研究对于智能电网安全运行具有重要意义。本文主要围绕虚假数据注入攻击的检测以及容侵控制展开研究，利用改进的 Conformer 模型进行虚假数据的定位检测；在此基础上，融合堆叠双向门控循环单元(Stacked Bidirectional Gated Recurrent Unit，SBIGRU)与混合注意力机制(Hybrid Attention Mechanism，HA)，探究了 SBIGRU-HA 模型的检测效果；同时结合自编码器(Autoencoder，AE)和带有梯度惩罚的生成对抗网络(Wasserstein Generative Adversarial Network with Gradient Penalty，WGANGP)，运用 AE-WGANGP 实现系统遭受攻击后所需的数据恢复。具体研究内容如下：

首先，开展了基于改进 Conformer 的虚假数据注入攻击检测研究。利用 Meta-Acon 激活函数取代卷积下采样原模块中 Relu 激活函数；在此基础上，在原 Conformer 模块中加入卷积模块，并更改模块顺序；同时，选择 SMU 激活函数替换 Conformer 算法中的 Swish 激活函数；随后在 IEEE-14 以及 IEEE-57 节点测试系统上进行了仿真验证，仿真结果表明，改进后模型的各项检测指标均有所上升。

其次，探讨了堆叠双向门控循环单元与混合注意力机制的检测模

型 SBIGRU-HA。采用 SBIGRU 提取给定时间段内的系统时序特征；同时，引入残差网络融合原始输入与 SBIGRU 捕获的时序特征；在此基础上，融合坐标注意力(Coordinate Attention，CA)、卷积注意力(Convolutional Block Attention Module，CBAM)、无参注意力 SimAM 三种注意力机制，提取数据的时空特征并为被注入攻击的特征分配更高的权重；在 IEEE-14、IEEE-57 节点测试系统上进行仿真实验，仿真结果表明其检测效果比改进后的 Conformer 更优秀，能有效的检测出虚假数据注入的位置。

最后，研究了虚假数据注入攻击后系统的容侵控制。生成两种数据集：一种包含随机单节点攻击，另一种包含随机多节点攻击。结合 AE 和生成对抗网络(Generative Adversarial Network，GAN)，并引入 Wasserstein 距离和梯度惩罚项以避免 GAN 在对抗训练中容易出现梯度消失和模式崩溃的问题；采用对抗训练，达到纳什平衡，学习数据间的空间关系，完成数据的重构；仿真结果表明，所采用的恢复方案能够准确地将被污染的状态量恢复到攻击前的值，对于 FDIA 污染具有较高的恢复精度。

综上，本文研究的检测算法以及容侵控制方案能有效实现对智能电网的 FDIA 定位检测以及数据恢复，为维护智能电网安全、稳定运行具有重要的理论研究意义与实际应用价值。

关键词：智能电网；虚假数据注入攻击；攻击检测；容侵控制；Conformer；生成对抗网络

Research on Attack Detection and Intrusion Tolerance Control of Smart Grid under False Data Injection

ABSTRACT

Smart grid is a modernized grid system using advanced technology to achieve efficient, safe, reliable, environmentally friendly and economic operation of the power system. However, false data injection attack (FDIA) is one of the most destructive malicious attacks against smart grid. By designing attack vectors to tamper with the measured data and avoid the bad data detection (BDD) of the state estimation in the energy management system (EMS), it causes the dispatching center to misjudge the current power grid state, thus achieving the purpose of destroying the security and stability of the power system or obtaining illegal benefits. Therefore, it is of great significance to carry out research on FDIA attack detection and tolerant intrusion control for the safe operation of smart grid. This paper mainly focuses on the detection of false data injection attacks and intrusion tolerance control. The improved Conformer model is used to detect false data; On this basis, stacked bidirectional gated recurrent unit (SBIGRU) and hybrid attention mechanism (HA) are combined to explore the detection effect of SBIGRU-HA model; Meanwhile, combined with autoencoder (AE) and wasserstein generative adversarial network with gradient penalty (WGANGP), AE-WGANGP is utilized to implement the data recovery after the system is attacked. Specific research contents are as follows:

Firstly, the detection of false data injection attack based on improved Conformer is studied. The relu activation function is replaced by Meta-Acon activation function in the original Convention_Subsampling

module; On this basis, the convolution module is added to the original Conformer module, and the module order is modified; At the same time, SMU activation function is selected to replace Swish activation function in Conformer algorithm. The simulation results on IEEE-14 and IEEE-57 bus test systems show that the detection indexes of the improved model are improved.

Secondly, the detection model of stacked bidirectional gated recurrent unit and hybrid attention mechanism SBIGRU-HA is discussed. SBIGRU is used to extract the system time series characteristics in a given time period; At the same time, the residual network is introduced to fuse the original input with the time series characteristics captured by SBIGRU; On this basis, three attention mechanisms, namely coordinate attention (CA), convolutional block attention module (CBAM) and SimAM, are fused to extract the spatio-temporal features of data and assign higher weights to the features injected with attacks. Simulation results on IEEE-14 and IEEE-57 bus test systems show that its detection effect is better than that of the improved Conformer, and it can effectively detect the position of false data injection.

Finally, the intrusion tolerance control of the system after false data injection attack is studied. Two kinds of data sets are generated: one contains random single-node attack and the other contains random multi-node attack. AE and generative adversarial network (GAN) are combined, and Wasserstein distance as well as gradient penalty term are introduced to avoid the gradient disappearance and pattern collapse of GAN in confrontation training; Confrontation training is adopted and reaches Nash equilibrium, which can learn the spatial relationship

between data and complete the reconstruction of data; Simulation results show that the restoration scheme can accurately restore the contaminated state variables to the values before the attack, and has high restoration accuracy for FDIA contamination.

To sum up, the detection algorithm and intrusion tolerance control scheme studied in this paper can effectively realize FDIA location detection and data recovery of smart grid, which has important theoretical research significance and practical application value for maintaining the safe and stable operation of smart grid.

Wang Ruiren(Electronic Information)
Supervised by Wei Lisheng

Key words: Smart grid; False data injection attack; Attack detection; Intrusion tolerance control; Conformer; Generative adversarial network

目　　录

插 图 清 单

插 表 清 单

第 1 章 绪论

1.1 选题背景及意义

近年来，由于环境问题以及煤炭、天然气和石油等不可再生能源的资源有限，势必要加大清洁能源的生产并提高利用效率。虽然水力、生物质能、太阳能等可再生能源储量丰富，但其产生的电力受季节、天气等因素影响较大，因此不能很好地实现持续稳定的供电。因此，必须采用更加先进的技术，以应对日益增长的能源需求，并确保电力供给的可靠性和安全性。这需要不断推动科技创新，探索更高效、更环保的能源生产和供应方式。如今，电网开始实现能源网与信息网的融合。分布式能源的集成化日益增加了电网的复杂性，由于分布式能源的间歇性特性，其需要一个持续的监测系统。这种复杂性的增加给电网带来了一些问题，因为它们需要适应经济、社会以及技术的要求，以确保电网的正常运行，并促进能源交易[1]。另一个复杂因素是电网中存在的能源生产者和消费者，比如说可再生能源和电动汽车。电网不仅需要通过“实时”通信系统对所有设备进行高效控制，并严格控制运行延时，而且还要创建冗余通信系统，通信系统还必须要具有安全性、可靠性和弹性等特性。这些因素都促使我们重新考虑如何以更高效、经济和环保的方式生产和利用电能。

智能电网是应对上述挑战的新范例[2]。其目标是有效地提供可持续、经济、可靠和安全的电力能源，并确保经济的持续发展以及环境的可持续使用[3]。智能电网是对 20 世纪传统电网改进而得到的，传统电网通常采用集中式发电，通过输电线路将电能传输至各个用户。而智能电网使用电力和信息的双向流动[4]，同时采用现代信息技术，能应对各种情况和事件。这涉及对电力需求、供给、传输和分配的动态管理，以应对各种突发事件或异常情况，以确保电网的安全运行和持续稳定的电力供应。这种动态管理需要及时调整电力资源的分配和利用，以满足不断变化的需求和情况，确保电网在各种情况下都能够有效地运行。具体而言，其可以被视为一个电力系统，它在发电、输电、变电、配电和消费方面以综合方式使用信息[5]。为了进一步提高智能电网的灵活性、安全性，以及能源供给与使用的可靠性，则必须重新思考电网与用户的互动问题，必须强化基础设施建设、加强信息通信技术与先进技术的融合发展。

信息通信技术(Information and Communication Technology，ICT)带来了与安全和隐私相关的额外问题。2010 年，Stuxnet 蠕虫病毒证明了网络安全的重要性，该病毒使用 u 盘部署，感染所有 Windows 计算机，而后攻击核电站工业控制系统，最后该病毒将离心机旋转到物理故障点，同时向控制系统提供错误的反馈[6]。2015 年乌克兰的电力数据采集与监控(Supervisory Control and Data Acquisition，SCADA)系统遭受网络攻击，攻击造成乌克兰大范围停电[7]。在此之后，2016 年乌克兰又一次遭受网络攻击，其首都附近一个配电站被关闭，影响了基辅北部以及周边地区电力输送，断电事件减少大约 200MW 电量输送，相当于基辅晚间能源消耗总量的五分之一。除此之外，美国等国家又多次发现了工业系统遭受到黑客入侵的实例，这引起了多国政府和学术界的高度重视[8]。

在智能电网中，网络攻击可能导致停电并提供误导性的监控信息，使控制系统产生不必要的行为，甚至导致物理系统故障。FDIA 是针对智能电网最具破坏性的攻击类型之一，由 Liu 等人于 2009 年首次提出[9]。FDIA 是通过设计攻击向量篡改系统的量测数据，并且能躲避系统检测，从而影响状态估计的结果，最终破坏电网安全稳定或者攫取不法收益[10]。FDIA 通过往系统中注入虚假数据，基于虚假的测量结果，电网的运行和控制将会改变，最终导致电网系统不稳定[11]；其对能源市场的危害也不容忽视，尤其是对区位边际价格(Locational Marginal Prices，LMP)的影响，这是因为区位边际价格依赖于准确的拓扑结构和实时测量数据，因此其中任何错误数据都会导致区位边际价格变化[12]。除攻击检测外对于那些已遭受攻击的电网，则需考虑如何保持其继续平稳运行。因此，针对 FDIA 的检测以及容侵控制研究对于智能电网安全、稳定运行具有重大意义。

1.2 国内外研究现状

FDIA 具有较高的隐蔽性和破坏性，它能够干扰电力系统状态估计过程，一个成功的 FDIA 不仅能篡改系统的量测值，而且还能躲避 BDD 的检测，通过影响状态估计器的输出，引起调度中心对当前电网状态的误判，最终对电力系统产生物理或经济影响。目前针对 FDIA 的研究可以分为三类：第一类主要研究如何对状态估计进行隐蔽攻击，篡改测量的同时躲避 BDD；第二类研究致力于提出各种行之有效的 FDIA 检测方法；第三类主要研究针对 FDIA 的容侵控制方法。接下来将分别阐述这三类研究的研究现状。

1.2.1 智能电网虚假数据注入攻击现状

对虚假数据注入攻击生成方法的研究主要分为以下三种情况：在部分电力信息未知时，如网络拓扑与支路的参数不能完全获取；在电网结构可控时，如窃取结构信息伪造拓扑结构；在有一定限制条件时，如通信或技术限制导致的仪表访问权限问题[13]。

在部分电力信息未知时构造虚假数据注入攻击，这种条件下相对而言更符合实际情况。文献[14]在不知道雅可比矩阵和状态变量分布假设的情况下，采用主成分分析近似方法研究了盲虚假数据注入攻击的一般问题，所提出的攻击被证明是近似隐形的。文献[15]利用交替方向法分离出攻击数据，然后对真实数据进行主成分分析转换求出近似拓扑矩阵，从而构造虚假数据注入攻击。文献[16]研究了配电系统中状态估计的 FDIA，攻击者可根据功率流或注入测量来近似系统状态。文献[17]研究了仅利用电力系统子网内信息设计的不可观察虚假数据注入攻击的物理后果，其假设攻击者只能在一定区域内使用包括拓扑、发电调度和负载信息在内的历史数据，并且可以修改测量值。结果表明，即使只有有限的信息，攻击者也可以利用它损害系统的可靠运行。文献[18]针对传统的稀疏攻击向量生成方案，将攻击区域内的所有总线作为变量，给出了在这种情况下仅考虑攻击总线为变量的攻击向量生成公式。文献[19]提出了一种基于矩阵重构和子空间估计的电网状态估计 FDIA 方法，该方法不仅克服了数据不丰富时测量噪声难以处理的缺点，而且它不还需要系统参数和拓扑信息，只需要有限周期的测量数据。文献[20]提出了一种架构，可以在不知道电力系统底层信息的情况下创建篡改测量向量来执行 FDIA。

在电网结构可控时构造虚假数据注入攻击。在这种类型的攻击中，假设攻击者不仅可以篡改电力系统采集到的测量数据，而且可以篡改反映电网拓扑结构中开关设备状态的离散数据[13]。文献[21]针对交流非线性电网模型，提出了一种基于几何方法的盲攻击，其无需电网拓扑以及输电线路的导纳，所提出的攻击对交流状态估计具有近似隐身性，对直流状态估计具有完全隐身性。文献[22]针对由自动电压控制器调节的电网，提出了一种新的 FDIA 策略。文献[23]提出了一种具有开发和探测机制的自适应FDIA方法，并且评估了该攻击对电力系统的影响。文献[24]对电力系统实时运行中的偶然性进行了分析，提出了一种新的虚假数据注入攻击方法。文献[25]研究了远程状态估计中的假数据注入攻击问题，将最优

攻击策略问题转化为可解的凸优化问题，从而得出具体需要攻击哪些传感器。文献[26]对局部协同网络物理攻击进行了相关研究，其仅利用不完整的网络信息，就能导致无法检测到的传输线路中断。文献[27]提出了一种网络拓扑攻击，能够在不威胁电力系统安全的前提下，干扰电力市场中的常规交易。文献[28]考虑了所有可能的网络拓扑攻击方法，即线路附加攻击、脱线攻击、线路切换攻击。针对不同的网络拓扑攻击场景建立了最优攻击模型，并使用元启发式优化算法—自然聚合算法来求解得到的攻击模型。

在有一定限制条件时构造虚假数据注入攻击。文献[29]研究了在多结算电力市场中，攻击者如何通过破坏传感器数据以达到从虚假的双边合同中获利的目的。文献[30]提出了一种只需要电力系统拓扑信息的零参数信息攻击。文献[31]揭示了数据攻击与物理后果之间的潜在联系，并分析了攻击者如何发起恶意数据攻击来触发连续中断，从而对电网造成巨大损害。文献[32]考虑到发电系统控制输入通道中的执行器饱和与 FDIA，分析攻击的行为和隐蔽性，设计了隐身攻击策略。文献[33]分析了虚假数据注入攻击对电力系统状态估计的物理后果，其提出了一种攻击框架，攻击者在该框架中通过对少量测量实现局部交流状态估计，从而满足系统的非线性交流特性。文献[34]讨论了攻击者在完全掌握电网拓扑结构的条件下构建稀疏攻击。文献[35]中采用机器学习技术进行 FDIA。具体来说，他们训练 GAN 来生成虚假的电力系统测量值。虽然文献[20]、文献[35]中都使用 GAN 完成 FDIA，但还是有一些区别。前者使用交流非线性潮流模型，后者采用直流线性潮流模型；前者的方法只需要正常测量，而后者需要正常和篡改测量来训练条件对抗网络（Conditional Generative Adversarial Network，CGAN）。

上述专家学者对虚假数据注入攻击进行研究并取得了一定的成果，但大部分文献仅考虑了直流情况，因此交流电网中的虚假数据注入攻击值得进一步探讨。

1.2.2 智能电网虚假数据注入攻击检测现状

近年来，许多专家学者针对虚假数据注入攻击检测进行了深度研究并有许多成果，大致可分为模型驱动和数据驱动两类[11]。

模型驱动又可分为基于估计的检测和其他直接计算方法。基于估计的 FDIA 检测方法通常是先状态估计或预测，将计算结果与这些状态的测量结果进行比较。而直接计算方法则依赖于系统的测量值和参数。

1）基于估计的检测方法：文献[36]用无迹卡尔曼滤波(Unscented Kalman Filter，

UKF)完成 FDIA 检测。考虑到当 UKF 更新步的量测量中存在虚假数据时，容易增大 FDIA 检测的误警率，文献[37]将基于极端梯度提升的日前负荷预测与 UKF 相结合，同时引入中心极限定理，实现对 FDIA 的精确检测和修正。文献[38]提出一种基于零序电压的三相配电系统 FDIA 检测方法，通过比较估计出的零序电压向量的 L2 范数与预定义阈值来检测 FDIA 的存在。文献[39]采用未知输入观察法完成检测，利用系统的内部状态，但增加了对系统模型的依赖。

2）关于直接计算的方法：文献[40]利用攻击特征的逻辑判断矩阵对 FDIA 的攻击点进行定位检测。文献[41]首次考虑了针对移相器的 FDIA，利用线路电流和节点注入电流与端电压的比值设计检测指标，在量测噪声和负荷波动场景下检测率依然很高。文献[42]提出一种检测方法用于检测线路电流差动继电器所遭受的 FDIA，其通过测量和计算两种不同的方法推导出 LCDR 终端处的叠加电压，并与阈值相比从而检测 FDIA。

数据驱动算法可以分为机器学习方法、数据挖掘方法和其他不涉及学习或挖掘的方法[11]。

1）基于机器学习的方法：文献[43]采用基于卷积神经网络的多标签分类器对 FDIA 进行定位，该分类器依赖于大量的训练数据集。文献[44]将异常检测问题转化为部分可观察马尔可夫决策过程问题，采用强化学习中的 SARSA 算法完成对 FDIA 的实时检测。文献[45]针对攻击者只知道部分网络拓扑信息的情况，提出一种攻击向量生成策略并且提出一种基于投票的集成学习技术来检测智能电网中的 FDIA。文献[46]将长短期记忆神经网络(Long Short Term Memory，LSTM)和 AE 相结合，该算法可以学习自动发电控制(Automatic Generation Control，AGC)系统的动态特性。用正常数据训练 LSTM-AE，检测阶段将数据输入 LSTM-AE 后比较残差完成检测，阈值由每个量测的最大重构误差决定。文献[47]应用卷积神经网络(Convolutional Neural Network，CNN)处理 FDIA 检测问题。为了模拟电网动态行为并捕获 FDIA 引起的异常数据特征，文献[48]提出采用循环神经网络检测 FDIA。文献[49]提出一种半监督深度学习方法，将自编码器集成到生成对抗网络中。文献[50]提出一种由污染状态分离方法、增强的坏数据识别方法和状态恢复算法组成的状态重建方案，完成对虚假数据的定位、剔除与修正。

2）基于数据挖掘的方法：其通过发现大数据集中的关联，从而得出关于数

据隐藏属性或模式的结论。文献[51]结合非嵌套广义样本(Non-Nested Generalized Exemplars，NNGEs)方法和状态提取理论(State Extraction Method，STEM)检测FDIA。文献[52]采用 Hoeffding 自适应树对 FDIA 进行分类，结果表明该分类器能适应网格负载的缓慢变化。文献[53]提出了一种顺序模式挖掘方法，准确提取电力系统干扰和网络攻击的模式。其展示了一种利用同步量测量进行电力系统干扰和网络攻击检测的方法，并强调了挖掘公共路径算法的前景。引入了术语“公共路径”。公共路径是指在时间顺序下的关键系统状态序列，其代表不同类型的干扰以及网络攻击。文献[54]首次将联邦学习应用于电网的攻击检测，提出一种将联邦学习、Transformer、Paillier 密码系统相结合的 FDIA 检测方法，其在保护数据隐私和减少通信开销方面比集中式检测方法更具优势。

3）其他不涉及学习或挖掘的方法：文献[55]将攻击矩阵范数小于特定阈值的行置零，以此保证解的收敛性，并与几种矩阵分离算法对比证明了所提算法的优越性。输电系统依赖监控指令，指令通道属于监视控制与数据采集系统，但针对它的攻击研究较少，文献[56]提出一种能够同时检测虚假数据和虚假命令注入攻击的通用方案，计算成本低，且不需要迭代。

针对虚假数据注入攻击检测，诸多专家学者不断研究，并取得了积极的进展，但电网测量数据不仅具有时间特性，同时还拥有空间相关性，因此在检测模型设计时充分考虑电网测量数据的时空特性具有重大意义。

1.2.3 智能电网虚假数据注入攻击下容侵控制现状

当前，诸多研究者也一直致力于不断更新 FDIA 容侵控制策略。文献[50]结合准牛顿和 Armijo 线搜索提出状态恢复方法，但为了获得高精度的恢复结果，需要尽可能多地成功识别受污染的测量值。文献[57]提出了一种针对直流微电网的弹性控制器，实现 FDIA 和拒绝服务攻击下的电压调节和电流共享。该方法结合电压和电流误差设计了开关二级控制器，并提出了一种基于增益的自适应控制方案，在控制参数设计中放宽了对网络攻击知识的要求。文献[58]提出了一种基于分布式滑模观测器的直流微电网二次控制方法。其设计了一个分布式滑模观测器来估计攻击信号，该观测器用于补偿控制输入，以消除 FDIA 信号对系统的不利影响。文献[59]提出了一种针对 FDIA 的分布式弹性二次控制方法，其通过设计自适应控制框架来减少攻击对系统的影响，保证了局部控制输入对攻击信号的弹性。文献[60]提出了一种自适应滑模控制器，其采用自适应机制在线估计攻击

信号的上界，并自动消除混合攻击的影响，从而保证了在攻击信息未知的情况下电力系统的可靠运行。文献[61]提出一种控制方案，该方案使网络控制系统能够检测和减轻 FDIA，同时补偿测量噪声和过程噪声。文献[62]提出了一种基于信任的协作控制器，可以减轻对通信链路攻击所造成的不利影响，其只需要本地和邻居的信息来检测和减轻攻击。文献[63]提出了一种利用并联 DC/DC 变换器减轻直流微电网中虚假数据注入攻击的方法，使用神经网络生成一个参考点给 PI 控制器，用 PI 控制器的输出减轻虚假数据对系统的影响。文献[64]提出了一种受拜占庭容错启发的数据弹性控制机制，其不需要对攻击者进行局部定位，但可以容忍任意的数据操作攻击，并收敛到最优值。文献[65]提出了一种针对智能电网中 FDIA 的检测与恢复机制，其采用改进的主成分分析法完成检测，并根据检测结果，设计了一种基于遗传优化的线性二次调节控制器来补偿 FDIA 对电压的变化。文献[66]结合 AE 和贝叶斯变化验证检测 FDIA，采用生成模型 AE-GAN 来生成包含正常操作条件下测量样本的多样化数据集，设计模式匹配算法根据曼哈顿距离从数据集中恢复伪造的测量值。文献[67]采用生成式对抗网络模型生成未篡改数据，用于恢复受损的测量数据。其需要大量正常和受攻击样本来训练高性能模型，但实际电网无法提供如此比例合适的数据集。文献[68]开发了一种基于 P-Q 分解的线性变换方法来重建受损状态变量。然而，潮流的非线性特性可能会降低恢复的精度。文献[69]提出了一种基于均值方差的优化方法来恢复仪表的值，但负荷的快速变化可能会影响恢复质量。

基于上述文献，尽管各种容侵控制方法能降低 FDIA 对电网的影响，但由于电网的运行对社会各个领域的正常运转都具有重要影响，如何进一步降低虚假数据注入攻击对系统的影响值得更深入地探索。

1.3 主要研究内容

本文主要探讨 FDIA 的检测及其容侵控制，首先利用改进的 Conformer 模型进行虚假数据的定位检测，其次融合 SBIGRU 与混合注意力机制，探究 SBIGRU-HA 模型的检测效果，最后结合 AE 和 WGANGP，采用 AE-WGANGP 完成系统遭受攻击后所需的数据恢复，全文的研究框架如图 1-1 所示，具体内容如下：

第 1 章：阐述了论文的选题背景和意义，分析了目前虚假数据注入攻击构建

的方法、攻击检测的方法和容侵控制研究的现状；

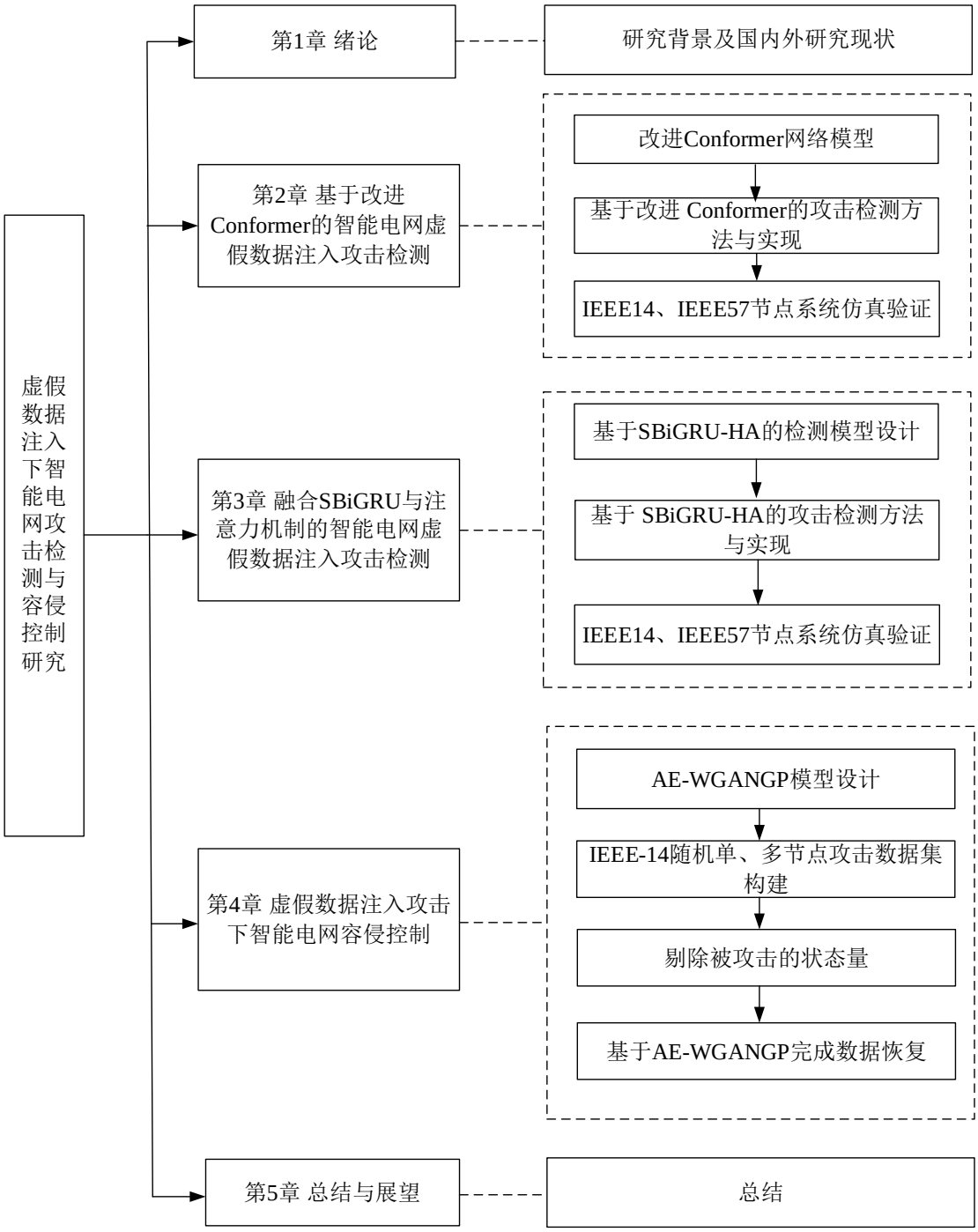


图 1-1 研究内容框架图

第 2 章：开展了基于改进 Conformer 的智能电网虚假数据注入攻击检测研究。阐述了 Conformer 模型的基本结构，包括卷积下采样模块、前馈模块、卷积模块以及多头自注意力模块，之后基于改进 Conformer 模型在 IEEE-14 以及 IEEE-57 节点系统进行实验，实验验证了算法的可行性和可靠性；

第 3 章：探讨了 SBIGRU 与混合注意力机制的检测模型 SBIGRU-HA。讨论

了堆叠双向门控循环单元的结构，针对 CA、CBAM 分别改进，并与 SimAM 注意力结合，最后基于 SBIGRU 与混合注意力机制的检测模型在 IEEE-14 以及 IEEE-57 节点系统分别开展实验，实验证明该算法比第二章的 Conformer 更优秀，更加胜任 FDIA 定位检测任务；

第 4 章：研究了虚假数据注入攻击后系统的容侵控制。探讨了自编码器、带梯度惩罚的 Wasserstein 生成对抗网络原理，阐述所采用 AE-WGANGP 中生成器、判别器模型具体结构以及对抗训练的具体流程。针对目前没有公开的 FDIA 数据集以及相应的攻击前的原始数据集，使用 MATPOWER 工具包在 IEEE-14 系统模拟生成两种数据集：一种包含随机单节点攻击，另一种包含随机多节点攻击。给出数据生成的流程图、系统攻击前后的残差以及攻击前后系统状态的变化，之后采用 AE-WGANGP 分别对系统的电压、相角进行状态恢复。数值结果表明，所采用的恢复方案能够准确地将被污染的状态量恢复到攻击前的值，对于 FDIA 污染具有较高的恢复精度；

第 5 章：总结全文主要研究内容，并分析下一步工作方向和展望。

1.4 本章小结

本章首先阐述课题的研究背景和意义；然后探讨虚假数据注入攻击的构建、虚假数据注入攻击检测及容侵控制的研究现状；最后概述全文的研究框架，并阐述各章的主要内容。

第 2 章 基于改进 Conformer 的智能电网虚假数据注入攻击检测

若仅实现针对虚假数据的检测，则不能精确地判断出哪些线路或状态受损。在实际应用中，除发现有无攻击外，准确地确定虚假数据注入攻击的位置，是迅速采取有效对策，并进行状态重构的关键。因而，将 FDIA 定位检测视为一个多标签二分类问题，采用改进的 Conformer 模块，用卷积来提取局部特征、利用自注意力机制来捕获全局表示，辨识系统量测量来达到对攻击的精确定位。在 IEEE-14、IEEE-57 节点进行实验验证，确认该方法检测并定位虚假数据攻击的可行性。

2.1 改进的 Conformer 网络模型

随着机器学习被广泛应用，机器学习逐渐与电网攻击检测结合，通过对量测数据进行训练学习及特征提取，挖掘 FDIA 的离群特性，进而实现检测。Conformer 模型具体结构图如图 2-1 所示：

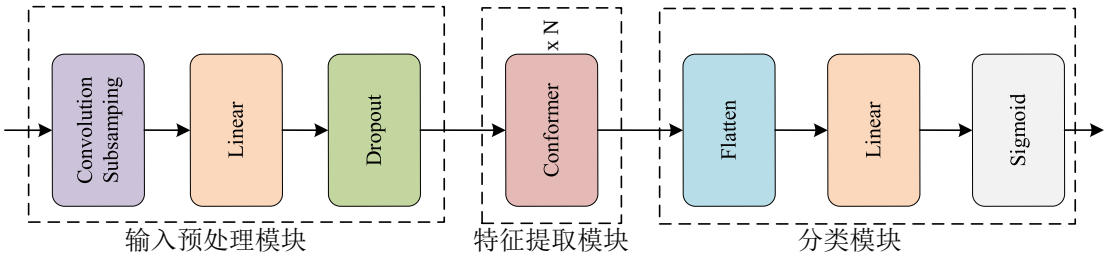


图 2-1 Conformer 网络结构

研究基于 Conformer 的网络结构，包括三个主要模块：输入预处理模块、特征提取模块以及分类模块。首先对序列数据进行特征提取，对其进行卷积下采样处理；其次通过 Conformer 模块用卷积来提取局部特征、利用自注意力机制来捕获全局表示。最后将各个层次的特征信息在全连接网络中进行融合，并通过 Sigmoid 激活函数完成分类预测。

2.1.1 输入预处理模块

输入预处理模块包含 Convolution Subsamping、Linear、Dropout 三个模块，而卷积下采样模块又由两个 Conv2d、Relu 交替串联而成，其中 Conv2d 的卷积核、步幅均为 1，填充为默认值，输入预处理模块具体结构如图 2-2 所示：

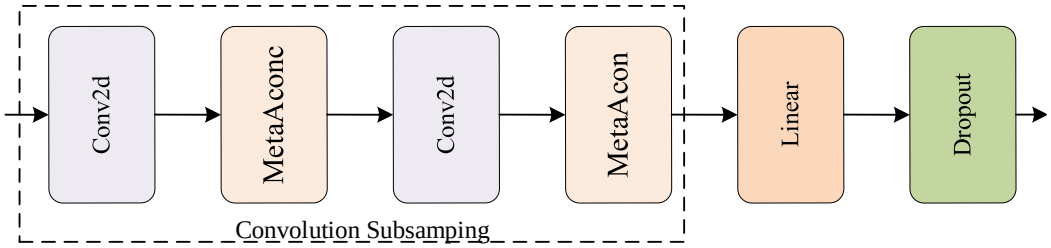


图 2-2 输入预处理模块网络结构

通过将激活函数函数引入到神经网络中，可以实现对非线性函数的任意近似，从而提高其表达能力。因此选用 Meta-Acon 函数替换原卷积下采样层中的 Relu 激活函数，自适应地决定每个神经元的激活状态，从而增强模型对数据特征的表达能力，提升学习效果和泛化能力。Meta-Acon 激活函数是在 ACON-C 的基础上修改得到的，通过定义开关因子 β 去学习线性映射和非线性映射之间的自适应切换[70]。Meta-Acon 函数表达式如下所示：

$$f(x)=(p_1-p_2)x\cdot\sigma[\beta(p_1-p_2)x]+p_2x \tag{2.1}$$

其中：$\beta=\sigma W_1W_2\sum_{h=1}^{H}\sum_{w=1}^{W}x_{c,h,w.}$；$x$ 表示输入的向量；p_1、p_2 均为神经网络待训练样本；β 为样本数据通道空间的自适应函数，不同通道内同一元素分配相同权值：σ 为 Sigmoid 函数；c、h、w 为空间维度尺寸；$W_1\in R^{C\times C/r}$,$W_2\in R^{C/r\times C}$；r 为缩放因子，通常为 2^n。

2.1.2 特征提取模块

特征提取模块结构如图 2-3 所示：

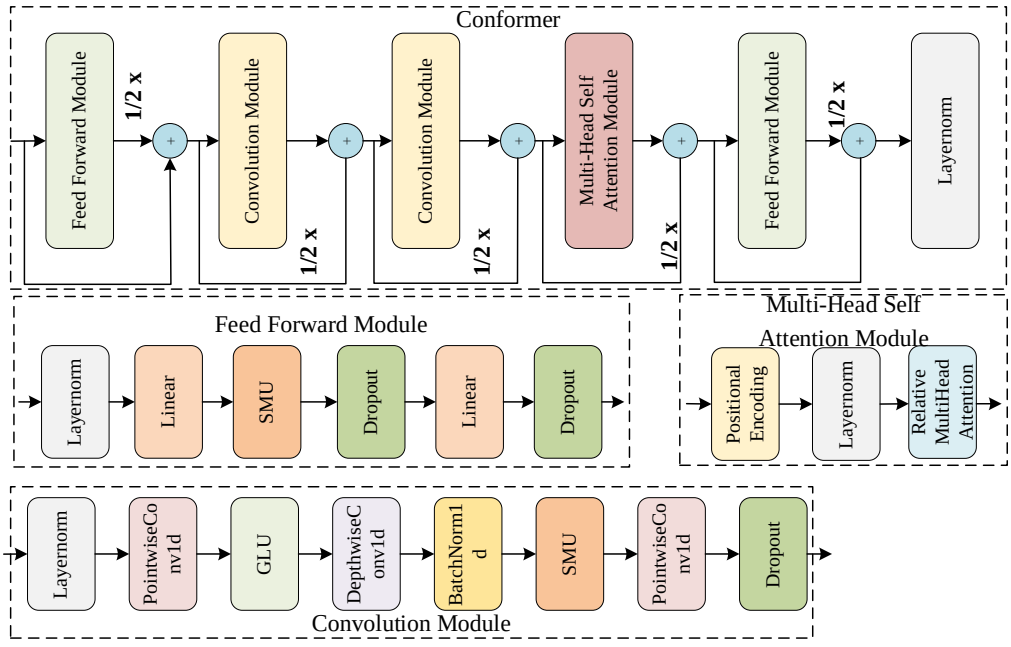


图 2-3 特征提取模块网络结构

由图可得，其是由两个前馈模块、一个多头自注意力模块、两个卷积模块以及一个层归一化组成。前馈模块由 Layernorm、Linear、Swish、Dropout 组成。卷积模块由 Layernorm、PointwiseConv1d、GLU、DepthwiseConv1d、BatchNorm1d、Swish、Dropout 组成。PointwiseConv1d 模块就是步幅为 1，填充为 0，卷积核为 1 的 Conv1d，而 DepthwiseConv1d 是步幅、填充都为 1，卷积核为 3 的 Conv1d。

改进后的 Conformer 模块在原 Conformer 中多增加一个卷积模块，并使用两个前馈网络模块将两个卷积网络模块和多头自注意力模块包围在内，序列数据先经由两个卷积模块再到多头自注意力模块，并在各模块之间做残差连接，提升卷积网络和多头自注意力模块在残差链接中的网络权重，增强网络的学习能力。将 FDIA 定位检测看作是一个多标签二分类问题，故将原模型中用于多分类的 Softmax 激活函数换为 Sigmoid 激活函数，完成非线性映射。

选用 SMU 函数替换原卷积和前馈模块中的 SWISH 函数。SMU 激活函数表达式如下：

$$f(x,ax;\mu)=\frac{1}{2}[(1+\alpha)x+(1-\alpha)x\cdot erf(\mu(1-\alpha)x)] \tag{2.2}$$

其中：$erf(\sigma)=\frac{2}{\sqrt{\pi}}\int_0^{\sigma}e^{-t^2}dt$，$erf(\cdot)$ 为高斯误差函数，α 为超参数，x 为网络模型中神经元的输入，μ 代表平滑参数，并为可学习的超参数。

图 2-4 中，超参数α和平滑参数μ分别设置为 0.01 和 0.3。

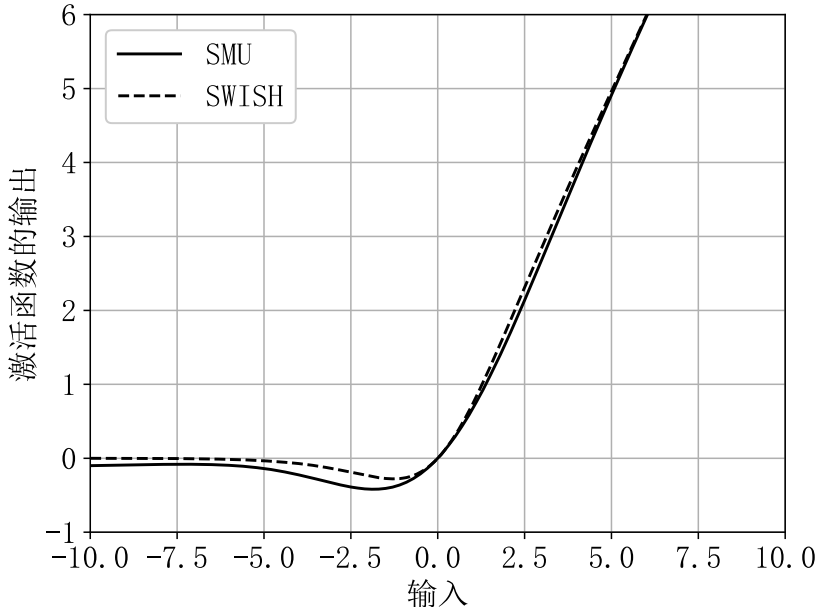


图 2-4 SMU 和 SWISH 函数曲线

在深度神经网络训练过程中，随着网络层数的增加，靠近输入层往往会导致

梯度弥散的现象的发生越来越加重[71]。由图 2-4 可知，SMU 激活函数无上界，可输出任意大的正值，因此可避免梯度消失的问题；同时因其连续可微性可避免出现急剧输出现象，有利于梯度的更新优化。相较于 SWISH 函数，SMU 还可以通过设置超参数α、自动调节平滑参数μ减少负值数据丢弃现象的出现，实现对深层神经网络的优化和性能的提升。

2.1.3 分类模块

分类模块就是由 Flatten、Linear、Sigmoid 组成。Flatten 负责将特征展开到一维，Linear 将特征线性映射到与检测点相同的数量，Sigmoid 将输出映射到零一之间，表示检测到攻击的概率。

2.2 实验验证及结果分析

2.2.1 数据集

首先从理论上而言，如果攻击者能够获得所有系统配置信息，并且有能力操纵所有仪表测量，那么很容易成功地启动 FDIA。在这种情况下，攻击者可以通过篡改系统的输入数据，例如修改测量值或传感器数据，来误导系统的运行并导致不良的控制决策。然而，在实际攻击中，攻击者不太可能已知系统全部信息。其次由于电力系统中 PMU 设备越来越多，基于纯 SCADA 测量的状态估计器正逐渐向混合状态估计器发展。对于攻击者而言，FDIA 模型应该进行相应的修改，否则控制中心就会利用 PMU 来检测网络攻击。实验所采用的数据集来自文献[50]。其设计的非线性攻击模型不仅考虑在有部分网络拓扑信息下的 FDIA，且还可同时处理 SCADA 和 PMU 的测量。最终分别在已知 IEEE-14 总线全部拓扑信息和部分 IEEE-57 拓扑结构信息的情况下实施 FDIA。数据以 15 分钟为间隔采集，共计 2480 个样本，其中正常数据 480 个，受损数据 2000 个。

2.2.2 评价指标及实验设置

检测领域常用的指标分别为准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F 分数(F_1)，以上指标值越大，代表分类器性能越好，其公式分别如下：

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2.3}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.4}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.5}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.6}$$

其中：TP 为真正类，表示模型正确地将正例预测为正例的数量；FP 为假正类，表示模型错误地将负例预测为正例的数量；TN 为真负类，表示模型正确地将负例预测为负例的数量；FN 为假负类，表示模型错误地将正例预测为负例的数量。因此，$TP + FN$ 等于数据集中受到攻击的样本总数，$TN + FP$ 等于数据集中正常数据的样本总数。

通过引入损失函数去寻找最优的学习参数集，并衡量每个小批的实际输出和真实输出之间的差异。选择交叉熵函数为损失函数以完成多标签分类。采用小批量梯度下降法对网络进行训练以提高收敛速度和避免过拟合。在实验中，每个小批量包含 256 个数据，按照机器学习惯例，从训练集中随机选择固定数量的训练样本来计算梯度，即一个 mini-batch。然后，将每批数据的 7/10 分为训练集，3/10 分为验证集。最后，使用 Adam 优化器，初始学习率设定为 0.001，patience 为 5。所有的测量在训练前首先使用 z-score 方法进行标准化，将每个量测量都标准化为平均值为 0，标准差为 1 的测量数据。实验平台环境为 Windows 11-X64 操作系统，显卡为 NVIDIA-GeForce-GTX-3060，16GB 运行内存，处理器为 12th Gen Intel(R) Core(TM) i5-12600KF 3.70 GHz。在 Pytorch1.10.2 框架、CUDA11.1 环境下进行实验。根据模型训练经验，选取多头注意力的头数为 10，前馈模块，多头自注意力模块，卷积模块的随机裁剪率均为 0.1，优化算法选取 Adam，epochs 设置为 200，每批数据 batch size 大小为 256，损失函数选取经典的交叉熵损失函数。采用原始 Conformer、GRU 模型作为对比实验。

2.2.3 实验结果与分析

1）IEEE-14

首先在已知所有拓扑信息并成功完成 FDIA 的 IEEE-14 节点进行 FDIA 检测。对每个状态量进行分类，0 为正常，1 为受到攻击，通过辨别每个状态量是否受到攻击来获取受攻击的精确位置，完成定位。其中量测量为 104 维，状态量为 28 维，量测数据为节点有功、无功，支路有功、无功，状态量为系统节点电压的幅值以及相角。Sigmoid 最终的输出为 0 到 1 之间，取 0.5 为阈值，小于 0.5

则认为为没被攻击，大于则认为被攻击。在 IEEE-14 节点系统进行仿真，得出每个状态量的 F 得分、准确率的结果，分别如图 2-5、2-6 所示：

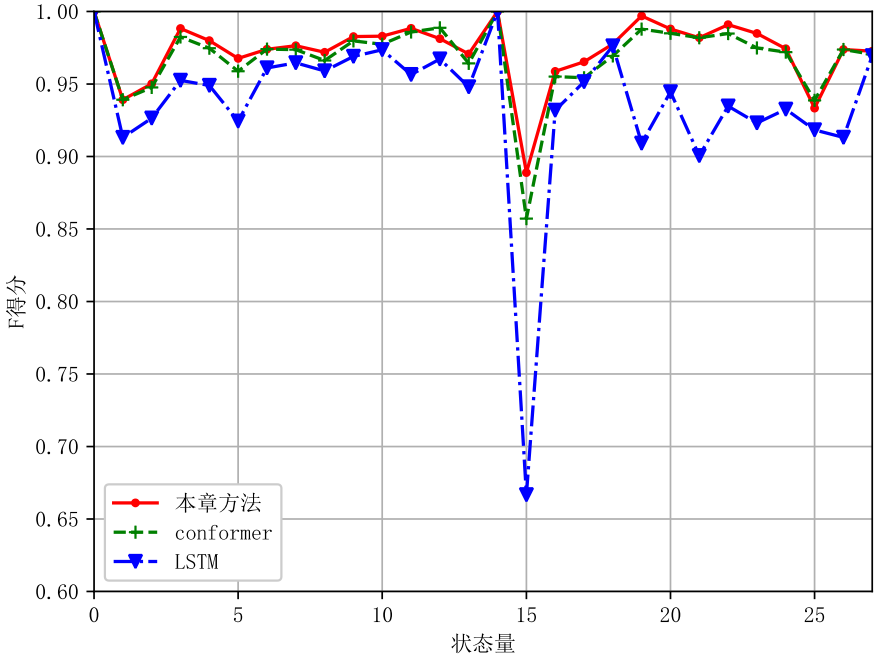


图 2-5 IEEE-14 节点系统各状态量检测的 F 分数

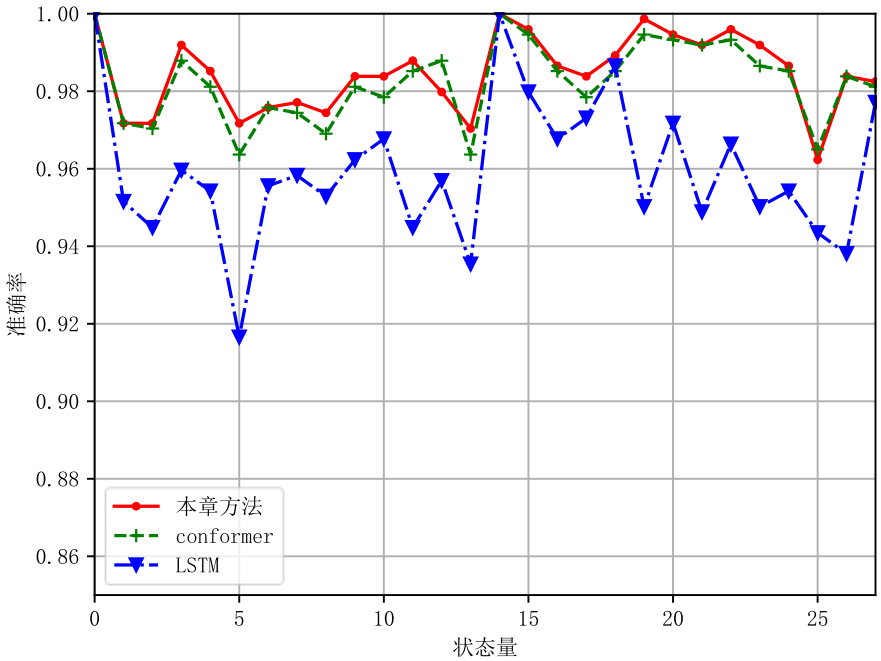


图 2-6 IEEE-14 节点系统各状态量检测的准确率

图 2-5 中横坐标按照先电压相角再电压幅值的顺序排列，其中 0 和 14 分别为平衡节点的电压相角和幅值。由图可知，LSTM 的 F 得分波动最大，最低点出现在 15，F 得分约为 66.7%，Conformer 稳定性次之，F 得分最低点也在 15，数值约为 85.7%，本章模型稳定性最优，最低点仅为 88.8%，也是出现在 15。三种

模型 F 得分最高点均为 100%，是因为平衡节点无法被攻击，其幅值和相角的标签值均为 0。通过利用 Meta-Acon、SMU 激活功能以及对 Conformer 模块的改进，如增加卷积模块、修改模块连接顺序和网络权重等，增强了网络的特征提取能力，获得了精度更高、泛化能力更强的网络模型。总体而言，改进后的 Conformer 基本在每个点的 F 得分都为三者中最高的，仅在 12 号状态量的 F 得分略低于 Conformer，其 F 得分分别为 98%、98.8%，仅略低 0.8%。综上所述，改进后的 Conformer 模型能更好的完成检测任务。

从图 2-6 可得出，本章模型检测准确率基本在每一个状态量均高于其他两种模型，仅在 12、15 号状态量低于 Conformer，在 12 号状态量的检测准确率相差最大，分别达到 98%、98.7%。本章模型检测准确率的最低值在 25 号状态量，约为 96.1%，Conformer 检测准确率的最低值在 13 号状态量，约为 96.3%。从稳定性而言，本章模型和 Conformer 不相上下，检测准确率均在 96%以上，LSTM 相对而言最不稳定，但准确率也都在 91%以上，检测准确率最低点出现在 5 号状态量，约为 91.6%。

描述正报率和误报率之间相对关系的接受者操作特征(Receiver operating characteristic，ROC)曲线如图 2-7 所示。ROC 曲线与坐标轴围成的面积记为 AUC(Area under curve)。

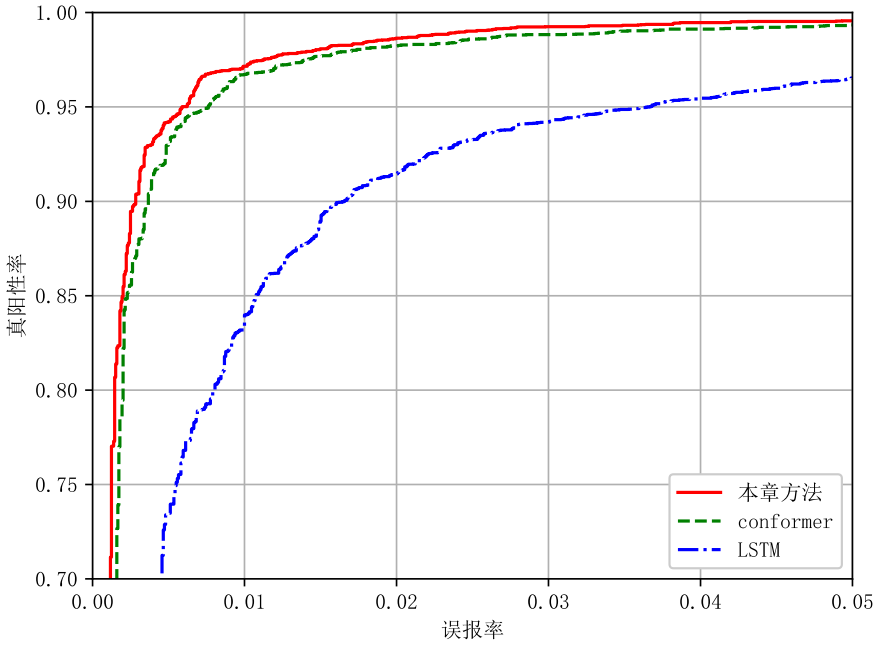


图 2-7 IEEE-14 节点系统 ROC 曲线

如果分类器表现优异，正报率将增加，导致 AUC 值接近 1。反之，若分类

器表现类似于随机猜测，则正报率将随误报率线性增加。ROC 曲线靠近左上角的程度与 AUC 值成正比，AUC 值越大表示模型性能越好。图 2-7 中本章模型、Conformer、LSTM 的 AUC 值分别为 0.9978、0.9971、0.9906。在较低的误报率情况下，改进后的模型也可以相对准确地识别受攻击的状态量。例如当误报率为 1%时，Conformer 和 LSTM 正报率分别约为 0.9670、0.8488，本章模型的正报率约为 0.9733，改进后的模型效果略胜一筹。

在 IEEE-14 系统上完成 FDIA 检测，各项检测指标如表 2-1 所示：

表 2-1 在 IEEE-14 节点系统的性能比较结果

结构	精确率	召回率	F 分数	准确率
LSTM	0.9558	0.9373	0.9465	0.9596
Conformer	0.9739	**0.9689**	0.9714	0.9825
本章模型	**0.9815**	0.9683	**0.9748**	**0.9846**

在表 2-1 中，可以清楚地看出，三种方法的所有指标均达到 90%以上。其中 LSTM 在所有指标上均最差，Conformer 次之，本章模型相对较好。其在精确率上分别比 LSTM、Conformer 高出 0.76%、2.57%；在召回率上比 LSTM 高出 3.1%，相较于 Conformer 略微下降 0.06%；在 F 分数上比 LSTM、Conformer 分别高出 2.83%、0.34%；最后在准确率上比 LSTM 高 2.5%，比 Conformer 高 0.21%。结果表明，所采用的检测方案对于 FDIA 检测并定位具有良好的性能。

2）IEEE-57

在已知部分拓扑信息并成功完成 FDIA 的 IEEE-57 节点进行 FDIA 检测。其中量测量为 419 维，状态量为 114 维。在 IEEE-57 节点系统进行仿真，得出每个状态量的 F 得分、准确率以及 ROC 曲线，分别如图 2-8、2-9、2-10 所示：

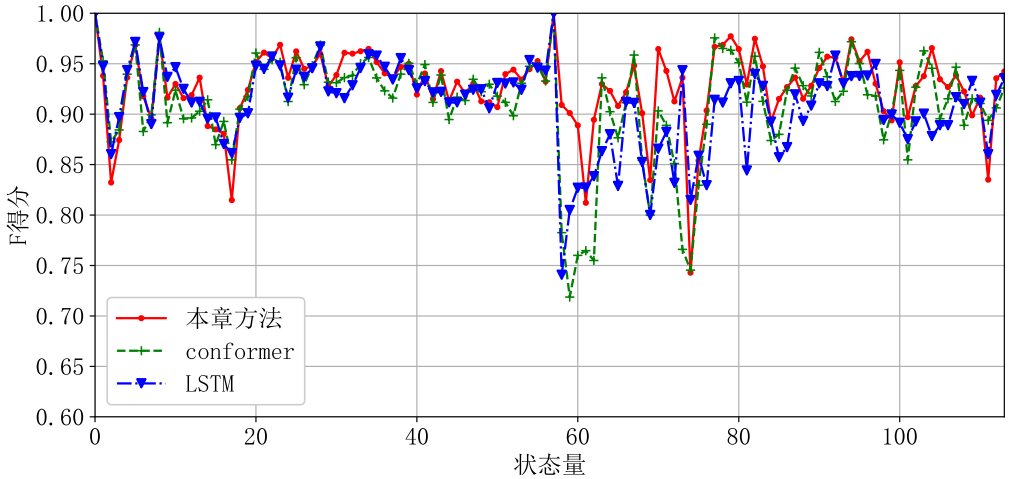


图 2-8 IEEE-57 节点系统各状态量检测的 F 分数

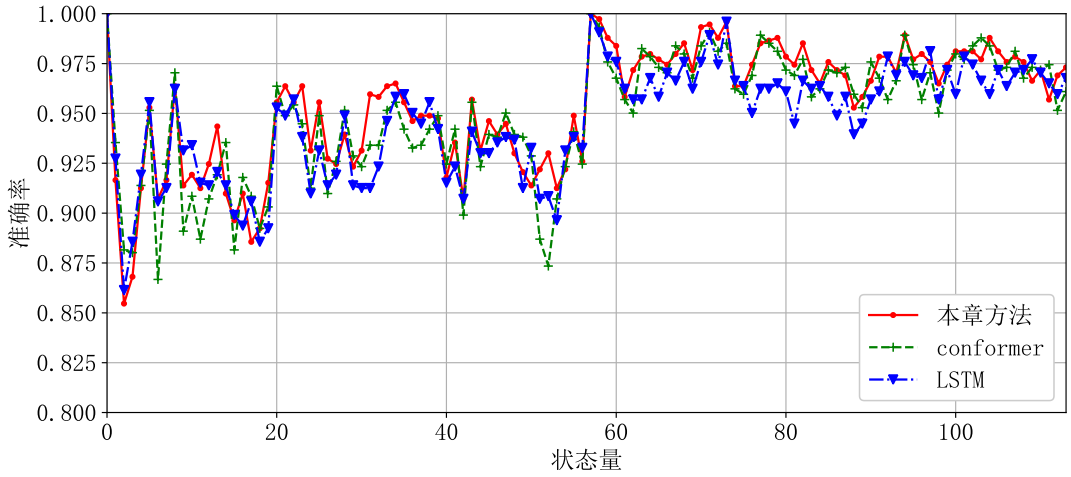


图 2-9 IEEE-57 节点系统各状态量检测的准确率

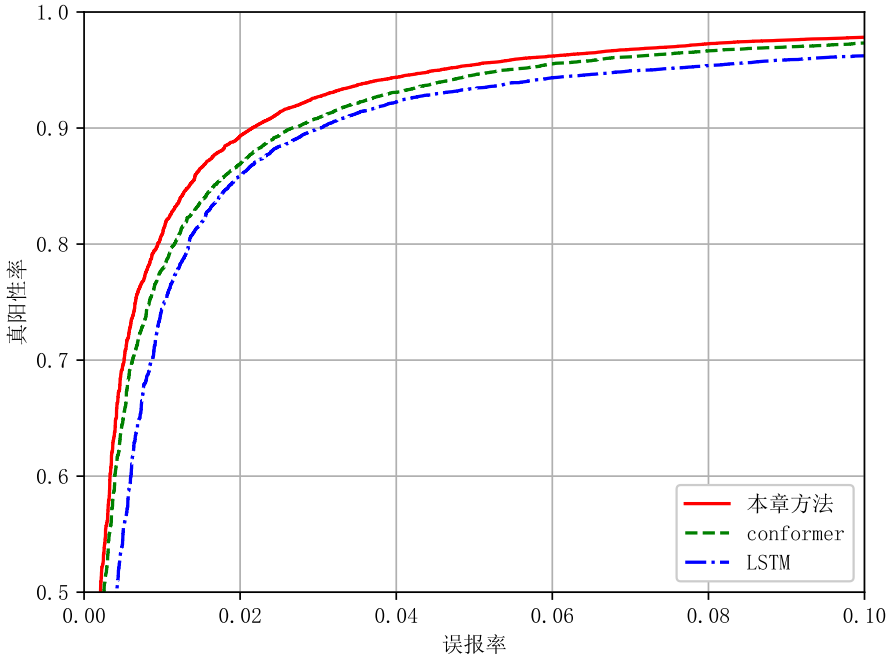


图 2-10 IEEE-57 节点系统 ROC 曲线

图 2-8 横坐标按照先电压相角再电压幅值的顺序排列，其中 0、57 号状态量分别为平衡节点的幅值和相角，攻击者要想成功完成 FDIA 并不被发现，则不能攻击平衡节点，所以 0、57 号状态量的标签均为 0。从稳定性而言，本章方法略优于 Conformer 以及 LSTM，LSTM 检测的 F 得分最低点在 58 号状态量，F 得分约为 73.9%，同样在这一点本章模型的 F 得分却能达到 90.7%，两者相差 16.8%；Conformer 检测的 F 得分最低点在 59 号状态量，F 得分约为 71.9%，而这一点本章模型的 F 得分却能达到 90%，两者相差 18.1%；本章模型检测的 F 得分最低点在 74 号状态量，F 得分约为 74%，而在这一点三者模型中 LSTM 效果最好，其 F 得分能达到 81.5%，两者也就相差 7.5%。

由图 2-9 可知，三种模型在各个状态量的检测准确率均在 85 以上，除去参考节点，本章模型检测效果最优的点出现在 58 号，检测结果达到 99.6%；Conformer 检测效果最优的点也出现在 58 号，检测效果略微低于本章模型，检测准确率约为 99.6%；LSTM 检测效果最优的点出现在 73，检测结果达到 99.5%；

随着系统节点数增加，数据维度急剧上升，三种模型的 AUC 值都分别有所下降。图 2-10 中本章模型、Conformer、LSTM 的 AUC 值分别为 0.9885、0.9868、0.9792。当误报率为 4%时，Conformer 和 LSTM 正报率分别约为 0.9309、0.9228，本章模型的正报率约为 0.9435，效果优于另两种模型。

在 IEEE-57 系统上完成攻击检测，各项检测指标如表 2-2 所示：

表 2-2 在 IEEE-57 节点系统的性能比较结果

结构	精确率	召回率	F 分数	准确率
LSTM	0.9311	0.9164	0.9237	0.9466
Conformer	0.9391	0.9113	0.9250	0.9494
本章模型	**0.9478**	**0.9182**	**0.9328**	**0.9542**

从表中可以看出，随着量测量的增加，数据维度变高，相较于低纬系统检测的各项指标均有所下降，但本章模型依旧在各项指标上都优于另外两种模型。其在精确率上分别比 LSTM、Conformer 高出 1.67%、0.87%；在召回率上分别高出 0.22%、0.69%；在 F 分数上分别高出 0.91%、0.78%；在检测准确率上分别高出 0.76%、0.48%。所提模型能很好地识别并定位受污染的状态量，防止漏检。

2.3 本章小结

开展了基于 Conformer 的智能电网虚假数据注入攻击检测研究。将攻击检测视为一种多标签分类问题，该算法与具体攻击模型及网络拓扑无关，通过收集历史数据，抽取样本特征，实现对异常数据的识别。在 IEEE-14 和 IEEE-57 系统进行大量实验，开展大规模仿真试验，验证所提方法的有效性，其具有更精确的定位检测性能。

第 3 章 融合 SBIGRU 与注意力机制的智能电网虚假数据注入攻击检测

当恶意数据被注入状态向量时，状态向量的时空数据相关性可能会偏离正常运行状态，传统的数据驱动检测算法无法捕获电网测量的时空相关性。因此，我们研究一种融合 SBIGRU 与混合注意力机制的检测模型 SBIGRU-HA。与以往的 FDIA 检测工作不同，SBIGRU-HA 不仅利用系统状态的空间数据特征来识别攻击，还从连续系统状态中呈现的时间数据相关性中进行学习。堆叠 BIGRU 用于捕获数据的时序特征，而 SBIGRU 则通过前向和后向 GRU 分别捕获输入数据的过去和未来信息。采用残差网络来融合原始输入和 SBIGRU 捕获的时序特征，并结合 CA、CBAM 和 SimAM 三种注意力机制。具体而言，CA 和 CBAM 并联，然后与 SimAM 串联，以便为被注入攻击的量测特征分配更高的权重。最后，将输出送入线性层和 Sigmoid 层，完成多标签分类检测。

3.1 融合 SBIGRU 与注意力机制的检测模型

随着机器学习被广泛应用，机器学习逐渐与电网攻击检测结合，其通过利用大数据及人工智能算法对电力 SCADA 系统中的量测数据进行训练学习及特征提取，挖掘 FDIA 的离群特性，进而实现检测[10]。SBIGRU-HA 模型具体结构图如图 3-1 所示：

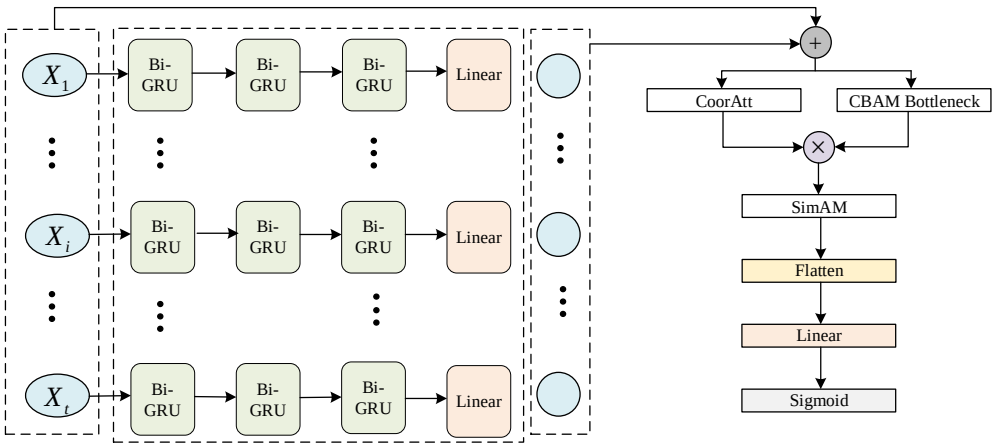


图 3-1 SBIGRU-HA 模型结构

SBIGRU-HA 模型主要分为两部分，前半部分为堆叠的 BIGRU 网络层，后半部分为融合后的注意力模块。BIGRU 依次串联组成 SBIGRU，而 BIGRU 分别

通过前向、后向 GRU 捕获时序输入数据的过去和未来信息。引入残差结构将原始输入与 SBIGRU 提取后的特征加权融合。SBIGRU-HA 中有三种注意力分别为坐标注意力、卷积注意力以及无参注意力 SimAM。融合后的特征表示分别送入坐标注意力和 CBAM 模块，再将两个模块的输出点积后送入无参注意力 SimAM 模块，依次通过扁平层、线性层以及 Sigmoid 非线性映射得到分类结果。

3.1.1 堆叠双向门控循环单元

GRU 将遗忘门和输入门合并到更新门中，从而直接在当前和历史状态之间创建线性依赖关系。它利用门机制可控的选择添加新信息和删除历史信息。其具体计算公式如下：

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{3.1}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{3.2}$$

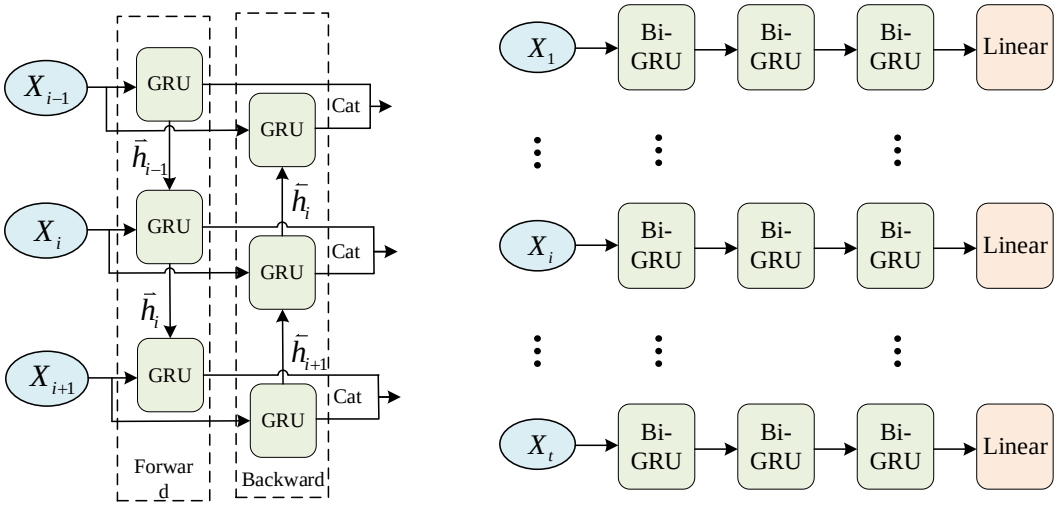
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{3.3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{3.4}$$

其中：x_t 为输入，z_t 为更新门，r_t 为重置门，σ 为 sigmoid 函数，可将数据变为 0-1 范围的数值，tanh 为双曲正切函数，可将数据变为[-1,1]范围的数值，h_{t-1} 为上一时刻的隐藏状态，$\tilde{h}_t$ 为候选隐藏状态，h_t 为传递到下一时刻的隐藏状态。

单向传播的缺点是不能利用未来的信息，特别是在处理长序列时，可能会遗漏一些重要的信息。此外，它容易受到梯度消失的影响，这可能导致模型的精度下降。在时间序列预测中，使用双向传播可以同时考虑过去和未来信息，从而更好地捕捉序列的过去和未来信息，提高模型的准确性。BIGRU 由正向 GRU 网络和反向 GRU 网络组成，其结合双向循环结构，以增加模型容量和灵活性。结构如图 3-2 所示，其中 X_i 代表 i 时刻的输入数据，i 时刻前向 GRU 同时接收 i 时刻的时序输入 X_i，以及上一时刻前向 GRU 的隐层输出 $\vec{h}_{i-1}$，后向 GRU 与此类似。

BIGRU 神经网络基于整个时间序列进行预测，它将隐藏层分为前向和后向两个相对的部分，分别读取过去和未来的时间信息。第一级前向 GRU 计算当前时间序列的序列信息，第二级后向 GRU 计算同一时间序列的逆序列信息，最后将两者拼接得到 BIGRU 的隐层输出。SBIGRU 由多个 BIGRU 堆叠而成。具体堆叠方法如图 3-3 所示。BIGRU 之间串行连接，最后一个 BIGRU 接线性层。



(a) BIGRU网络结构　　(b) SBIGRU网络结构

图 3-2 BIGRU 和 SBIGRU 网络结构

3.1.2 混合注意力机制

1）坐标注意力的改进

近年来，注意力机制广泛应用于机器学习任务中。其中，最经典的注意力是SE(Squeeze-and-Excitation)。而 SE 忽略位置信息，位置信息有助于生成空间特征图。CA 通过池化操作生成两个独立的方向感知特征图，然后对这两个特征图编码，生成两个注意力图，以捕获每个空间方向的依赖关系[72]。为此，我们采用改进后的坐标注意力，最终的输出不再是原始输入与沿水平和垂直方向注意力特征图的积，而是先将原始输入通过相应的卷积、BatchNorm 以及 Swish 非线性变换来获取增强后的特征表示，然后再与沿水平和垂直方向注意力特征图的积，从而在捕获通道信息和位置关系的同时提升网络的拟合效果。

坐标注意力结构如图 3-3 所示，其步骤如下：

首先，对于给定输入 X ，沿水平和垂直方向编码，其池化核分别为(H，1)，(1，W)。高度为 h 的第 c 个通道的输出 $Z_c^h(h)$ 为：

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i\le W}X_c(h,i) \tag{3.5}$$

相似的宽度为 w 的通道 c 的输出 $Z_c^w(w)$ 为：

$$Z_c^w(w)=\frac{1}{H}\sum_{0\le j\le H}X_c(j,w) \tag{3.6}$$

公式(3.5)和(3.6)分别为图 3-3 中的 H Avg Pool 和 W Avg Pool。上述两种变换分别对两个空间方向进行分解，得到一对方向感知特征图，这种转换使注意力能

够捕捉到沿某一空间方向的长距离依赖和沿另一空间方向的精确信息[72]。

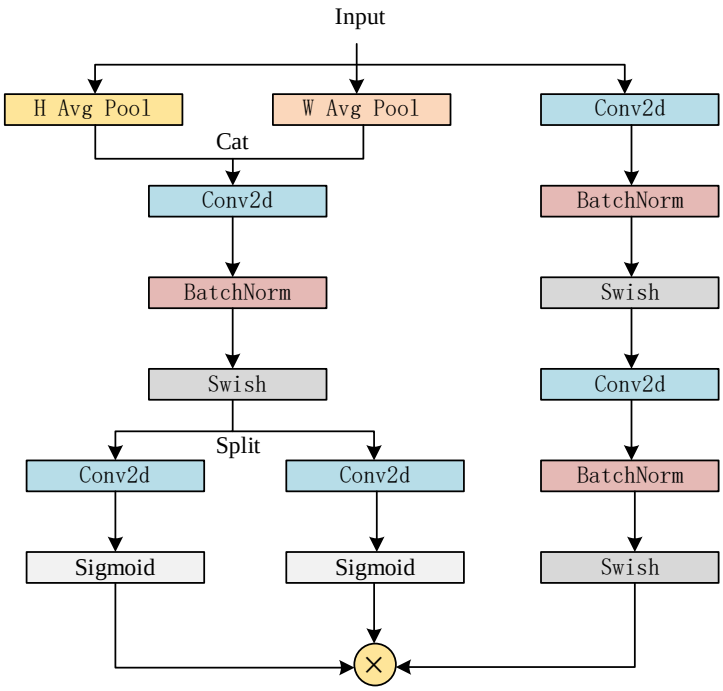


图 3-3 坐标注意力模型架构

其次，将沿两个方向的编码信息拼接后经过 1x1 的卷积、BatchNorm、Swish 函数运算得到中间特征映射 f。公式如下：

$$f = Swish(conv([z^h, z^w])) \tag{3.7}$$

其中：$f \in R^{C/r\times(H+W)}$，r 表示衰减率，方括号[,]表示沿空间维度的拼接操作，$conv$ 为 1x1 卷积，$Swish$ 为激活函数。

再次，将 f 沿空间维度方向分解，通过 1x1 卷积得到与输入 X 相同的维数后，再经过 $Sigmoid$ 得到两个方向上的注意权值：

$$g_c^h = Sigmoid(conv(f^h)) \tag{3.8}$$

$$g_c^w = Sigmoid(conv(f^w)) \tag{3.9}$$

其中：$conv$ 是 1x1 卷积，$Sigmoid$ 为非线性激活函数。

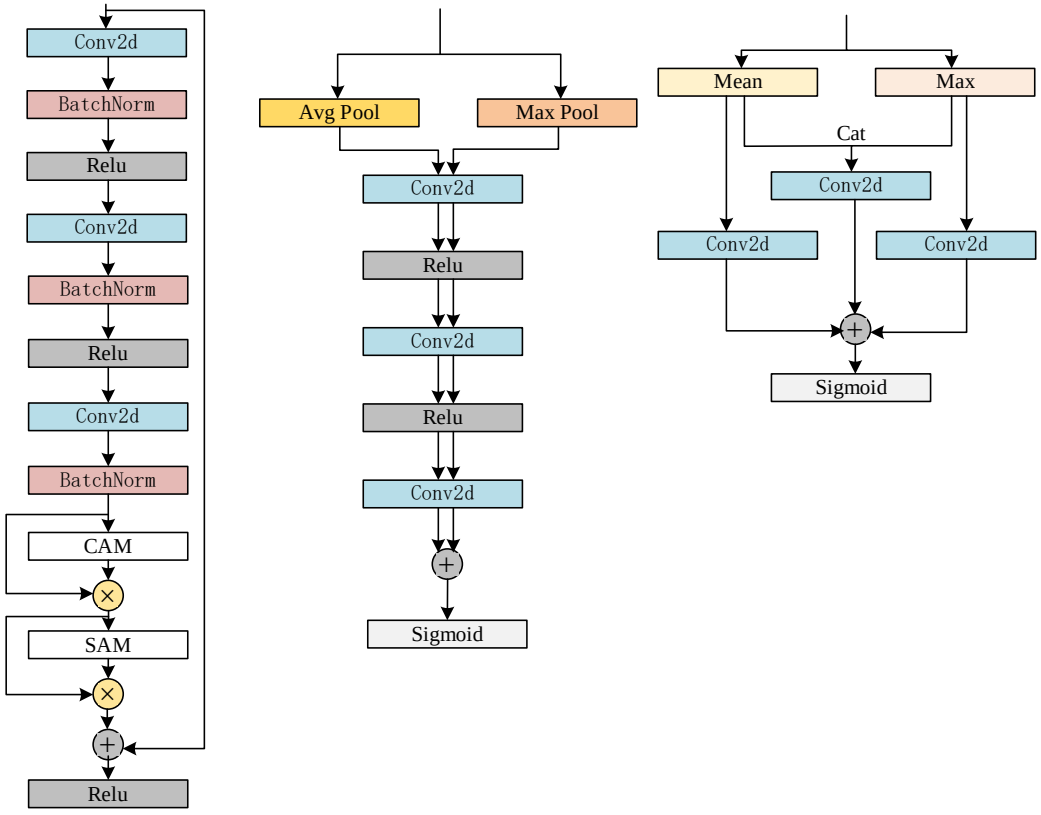
最后，将原始输入经过两层 conv2d、BatchNorm、Swish 后，与上述两个方向的权值点乘得到最终结果。公式如下：

$$y_c(i,j) = K(x_c(i,j))) \otimes g_c^h(i) \otimes g_c^w(j) \tag{3.10}$$

其中：$0 \le i < W$，$0 \le j < H$，g_c^h、g_c^w 分别表示在水平和垂直方向上的坐标注意力加权，K 表示两层 conv 1x1、BatchNorm、Swish 运算，$\otimes$ 表示点积。

2）CBAM 的改进

CBAM 是一个轻量级的注意力模块，包含两个子模块，通道注意模块(Channel Attention Module，CAM)和空间注意模块(Spatial Attention Module，SAM)。CBAM 整体网络结构如图 3-4 中(a)所示。对于 CBAM 的改动主要在其两个子模块，CAM、SAM 网络结构分别如图 3-4 中(b)、(c)所示。CAM 中在平均池化和最大池化后多添加一个卷积和 Relu 激活函数。SAM 的输出为原输入分别经过平均池化和最大池化后再分别卷积，并加上融合平均池化和最大池化再卷积的特征，最终再通过 Sigmoid 完成非线性映射。



(a) CBAM　　(b) CAM　　(c) SAM

图 3-4 CBAM 架构

CBAM 具体计算过程如下。首先，将 F 作为输入，其通过多层感知机(Multi-Layer Perceptron，MLP)后得到中间特征图 F'

$$F' = MLP(F) \tag{3.11}$$

其中：$F \in R^{C \times H \times W}$，$MLP$ 是指多层感知机。

其次，将中间特征图 F' 输入通道注意力模块，F' 分别经最大池化和平均池化后通过同一个 MLP' 后相加，再通过 Sigmoid 完成非线性映射，其输出与中间

特征图F'点积后得到F''

$$F'' = F' \otimes Sigmoid(MLP'(AvgPool(F')) + MLP'(MaxPool(F'))) \quad (3.12)$$

其中：MLP'表示 SAM 中的多层感知机，$AvgPool$、$MaxPool$分别表示平均池化、最大池化，$Sigmoid$为非线性激活函数，$\otimes$表示对应张量点积。

再次，将F''注入空间注意力模块，F''分别经平均池化和最大池化后再卷积，其输出与F''经平均池化和最大池化拼接后再卷积的结果相加，再通过非线性激活函数 Sigmoid 完成非线性映射，将其与F''点积后得到F'''

$$\begin{aligned} F''' = F'' \otimes Sigmoid(&f^{7\times7}(AvgPool(F'')) + f^{7\times7}(MaxPool(F'')) \\ &+ f^{7\times7}([AvgPool(F''); MaxPool(F'')])) \end{aligned} \quad (3.13)$$

其中：$f^{7\times7}$为卷积，7×7为卷积核尺寸，方括号[,]表示沿空间维度的拼接操作，$AvgPool$、$MaxPool$分别表示平均池化、最大池化，$Sigmoid$为非线性激活函数，$\otimes$表示点积。

最终将原始输入与F'''加权后送入 Relu 得到最终输出F''''

$$F'''' = Relu(F + F''') \quad (3.14)$$

3）融合后的注意力

通道注意力和空间注意力关注一维和二维的关系，而 SimAM 注意力通过设计能量函数直接生成 3D 权重，其无需额外的子网络和模型参数。SimAM 注意力使用能量函数 E 来计算目标像素点与周围像素点之间的关系，能量函数E计算公式如下：

$$E = \frac{4(\sigma^2 + \lambda)}{(t-\mu)^2 + 2 + 2\lambda} \quad (3.15)$$

其中：t为目标神经元，λ为常数，μ和σ^2为该通道内移除的目标神经元的均值和方差。

通过添加 Sigmoid 函数来抑制注意权值的离群值，并与输入特征矩阵的相应元素进行点积运算得到 SimAM 模块的最终输出$\bar{W}$为：

$$\bar{W} = Sigmoid(\frac{1}{E}) \otimes W \quad (3.16)$$

其中：W为 SimAM 模块输入，$Sigmoid$为非线性函数，$\otimes$表示点积，E表示能量函数。

最终将三种注意力相融合，改进后的 CA 和 CBAM 模块并行，两者输出结果点积后输入 SimAM 模块。在 CA 和 CBAM 后加入 SimAM 注意机制，其通过评估每个神经元的重要性来调整特征图的注意分布，使通道和空间注意协同工作。同时，SimAM 可以自适应调整特征映射的权值，更加关注目标的局部区域。这样可以提高目标定位精度，减少定位误差，增强网络的特征提取能力。

3.2 实验验证及结果分析

从机器学习的角度来看，位置检测问题就是一个多标签二分类问题，即将每个量测量进行分类，0 表示该量测量正常，1 表示受到污染，通过辨别每个量测量是否受到污染来获取受攻击的精确位置，进行定位检测。所采用数据集与上一章中使用的相同，采用 LSTM、GRU、改进后的 Conformer 模型在相同工况下进行对比实验，实验结果分别记录为 LSTM、GRU、Conformer-I。

1）IEEE-14

在详细掌握了系统的所有拓扑信息之后，通过构建攻击向量，对 IEEE-14 节点系统发起攻击并检测。针对虚假数据注入攻击的定位检测实际上就是对系统中的每个状态量进行分类，分类最终有两种结果，0 表示该状态量正常，1 表示受到攻击，检测完成后以 0.5 为界，小于 0.5 的都认为没有被攻击，大于 0.5 的都认为被攻击。比较 SBIGRU-HA 在不同 BIGRU 层数下的得到的检测指标，检测结果如表 3-1 所示：

表 3-1 在 IEEE-14 节点系统的性能比较结果

结构		精确率	召回率	F 分数	准确率
GRU		0.9536	0.9397	0.9466	0.9580
LSTM		0.9557	0.9373	0.9464	0.9596
Conformer-I		0.9815	0.9683	0.9748	0.9846
本章模型	1 层 BIGRU	0.9846	0.9782	0.9814	0.9856
	2 层 BIGRU	0.9854	0.9795	0.9825	0.9865
	3 层 BIGRU	**0.9868**	**0.9797**	**0.9832**	**0.9870**

由表 3-1 中可得，当 BIGRU 的层数从一层增加到三层时，四个指标也都略有增加，因此后续实验选择的 BIGRU 层数都为三。显然 SBIGRU-HA 的检测结果在精确率、召回率、F1 分数、准确率四个指标上均优于传统的 GRU、LSTM 以及优化后的 Conformer。以精确率而言，SBIGRU-HA 更是达到 98.68%，比传

统 GRU、LSTM、Conformer-I 分别高出 3.32%、3.11%、0.53%，召回率则分别高出 4%、4.24%、1.14%。在电网攻击检测中，漏检测比误检测的危害性更大，因此评价指标 F 得分则具有重要意义。而 SBIGRU-HA 的综合指标 F 分数则分别比传统的 GRU、LSTM 以及优化后的 Conformer 分别高 3.66%、3.68%、0.84%，准确率则又相比 GRU、LSTM、Conformer-I 分别高 2.9%、2.74%、0.24%。相对传统 GRU、LSTM，优化后的 Conformer 在低维系统检测效果和 SBIGRU-HA 相差不大。综上所述，当 SBIGRU-HA 中有三层 BIGRU 时，检测的各项性能指标最优，能更好地完成针对虚假数据的检测任务。

在 IEEE-14 节点系统分别采用 GRU、LSTM 以及有三层 BIGRU 的 SBIGRU-HA 模型进行虚假数据检测，得出每个状态量的 F 得分、准确率的结果，分别如图 3-5、3-6 所示：

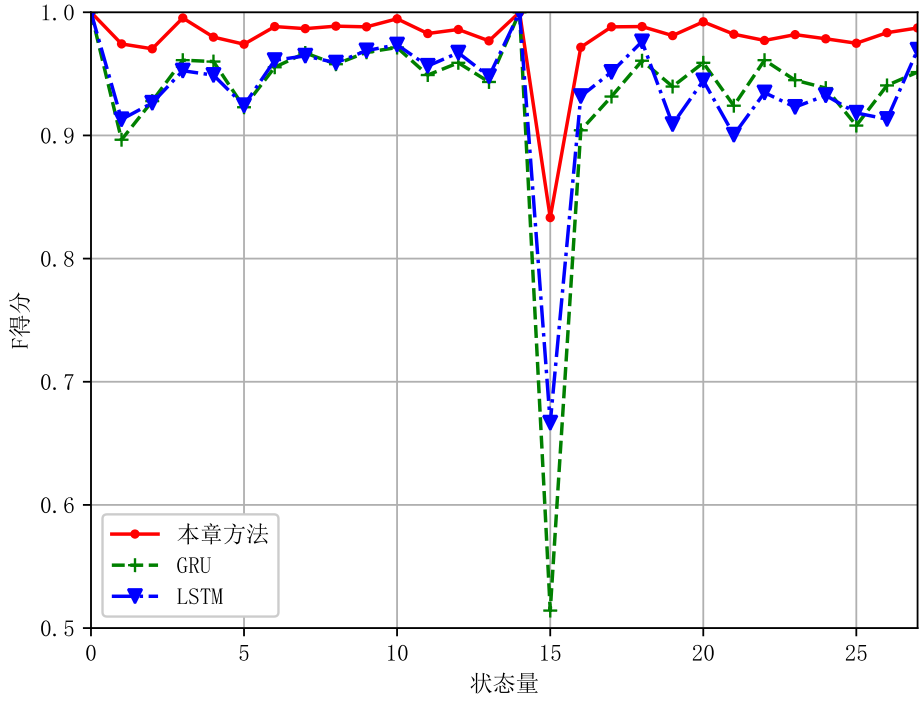


图 3-5 IEEE-14 节点系统各状态量检测的 F 分数

由图 3-5 可得，SBIGRU-HA 在各个状态量的 F 得分均优于传统的 LSTM、GRU，相对而言 GRU 的效果最差，其中相差最大的是一号和十五号状态量。一号状态量的检测 F 得分从原先的 89.5%上升到 97.3%，相比于 LSTM 的结果提升 7.8%；而十五号状态量的检测 F 得分由原先的 51.6%提升到 83.3%，相比于 LSTM 的结果足足提升 31.7%。从稳定性而言，LSTM 和 GRU 在各个状态量的 F 得分波动较大，其中 GRU 的最低 F 得分仅为 51.5%，而 LSTM 的最低 F 得分为 66.9%，

最高准确率都能达到 100%，而 SBIGRU-HA 最低 F 得分却能达到 83.3%，最高准确率也能达到 100%，明显稳定于上述两种传统模型。图中 0 和 14 号状态量检测的 F 得分之所以能达到 100%，是因为平衡节点无法被攻击，其标签值均为 0。总体而言，在 F 得分这一指标上，SBIGRU-HA 的检测效果更优秀。

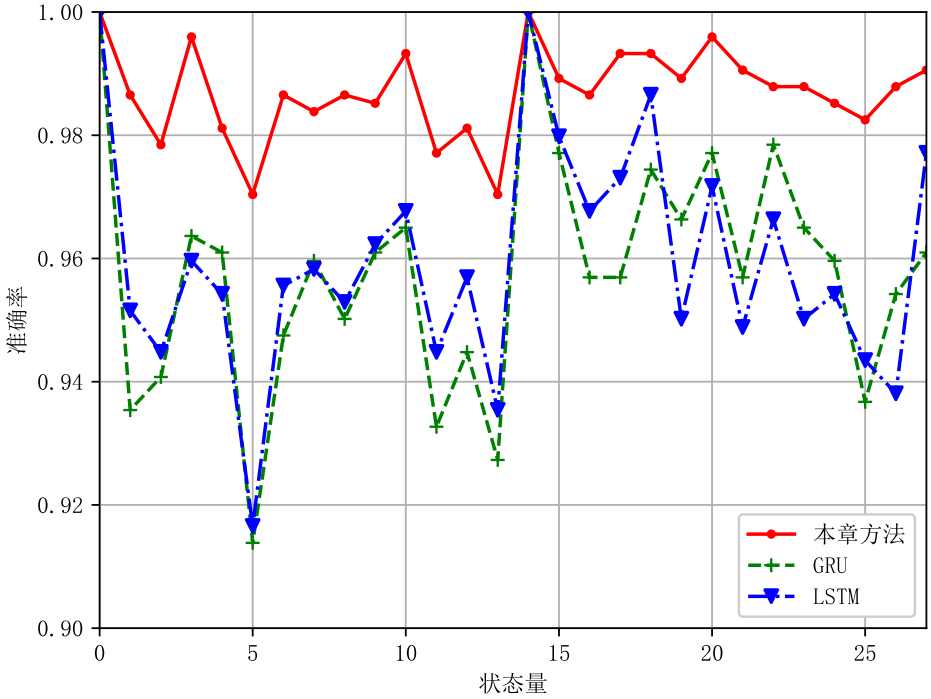


图 3-6 IEEE-14 节点系统各状态量检测的准确率

从图 3-6 中可以得出，SBIGRU-HA 准确率明显优于 LSTM、GRU。具体而言，LSTM、GRU 的最低检测准确率出现在状态量 5 号，仅为 91.3 和 91.6，而在此点 SBIGRU-HA 却能达到 97%。除去参考节点，LSTM 的最高检测准确率出现在状态量 18 号，达到 98.6%，而在此点 SBIGRU-HA 却依旧达到 99.3%；GRU 的最高检测准确率出现在状态量 22 号，达到 97.8%，SBIGRU-HA 结果为 98.8%。SBIGRU-HA 在每个状态量的检测准确率全部优于 LSTM、GRU。从稳定性而言，SBIGRU-HA 仅有 5 号和 13 号状态量的检测率为 97%，其他状态量的检测准确率均在 97%以上。GRU 和 LSTM 的稳定性相对而言较差，其检测准确率的最低点都出现在状态量 5 号，最低准确率分别为 91.4%、91.7%。换而言之，GRU 和 LSTM 的检测结果可信度不及 SBIGRU-HA 提供的高。这表明 SBIGRU-HA 在检测任务中具有更高的准确性和可靠性。

在 IEEE-14 节点系统分别使用 GRU、LSTM 以及有三层 BIGRU 的 SBIGRU-HA 完成对虚假数据的检测，其 ROC 曲线如图 3-7 所示：

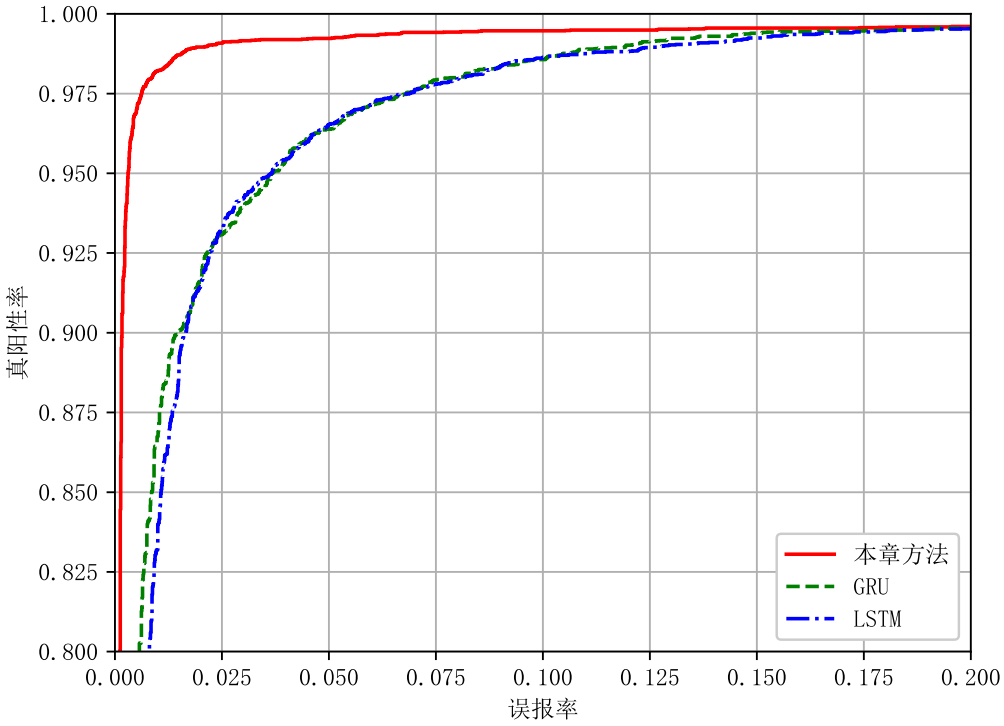


图 3-7 IEEE-14 节点系统 ROC 曲线

图 3-7 中 SBIGRU-HA、GRU、LSTM 的 AUC 值分别为 0.9966、0.9920、0.9906。当误报率为 1%时，GRU 和 LSTM 正报率相似，约为 0.9867，SBIGRU-HA 的正报率约为 0.9948。

2）IEEE-57

在已知部分拓扑信息的情况下，对那些成功注入 IEEE-57 节点系统的虚假数据进行检测。其中量测量为 419 维，状态量为 114 维，量测数据也为节点有功、节点无功，支路有功、支路无功，状态量为系统节点电压的幅值以及相角。比较 SBIGRU-HA 在不同 BIGRU 层数下的得到的检测指标，结果如表 3-2 所示。

表 3-2 在 IEEE-57 节点系统的性能比较结果

结构		精确率	召回率	F 分数	准确率
GRU		0.9358	0.9218	0.9287	0.9512
LSTM		0.9311	0.9164	0.9237	0.9466
Conformer-I		0.9478	0.9182	0.9328	0.9542
本章模型	1 层 BIGRU	0.9605	0.9433	0.9518	0.9665
	2 层 BIGRU	0.9615	0.9439	0.9526	0.9670
	3 层 BIGRU	**0.9636**	**0.9443**	**0.9539**	**0.9679**

从表 3-2 中可以看出，BIGRU 的层数从一层增加到三层时，精确率、召回率、F 分数、准确率都略有增加。因此后续实验中选择 BIGRU 层数为三层。当

数据维度上升，SBIGRU-HA 以及优化后的 Conformer 模型检测的各项指标相对低维都有所下降，但其与传统 GRU、LSTM 的模型依旧有提升。具体而言，SBIGRU-HA 精确率分别相对 GRU、LSTM、优化后的 Conformer 分别提高 2.78%、3.25%、1.58%，召回率分别提高 2.25%、2.79%、2.61%，F 得分分别提高 2.52%、3.02%、2.11%，准确率分别提高 1.67%、2.13%、1.37%。总体而言，本章模型 SBIGRU-HA 能更准确地检测出系统中的虚假数据。

分别采用 GRU、LSTM 以及有三层 BIGRU 的 SBIGRU-HA 模型，在 IEEE-57 节点系统进行仿真实验，得出每个状态量的 F 得分、准确率的结果，分别如图 3-8、3-9 所示：

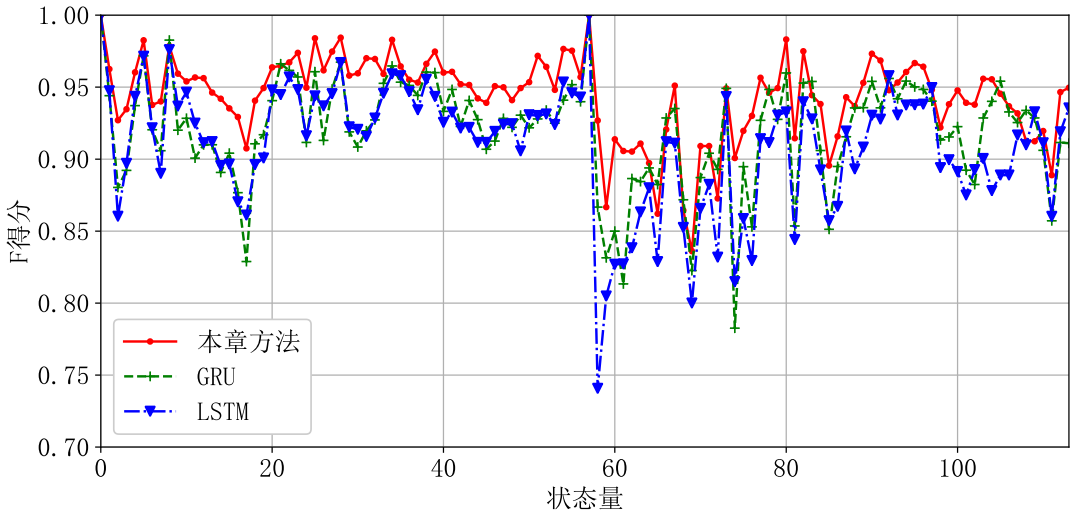


图 3-8 IEEE-57 节点系统各状态量检测的 F 分数

由图 3-8 可知，整体而言 SBIGRU-HA 在各个状态量的 F 得分优于传统 GRU、LSTM，但有部分点的检测效果差于它们，比如 83、92、97、105、108、109 号状态量，其中有明显的差距的是 108、109。具体而言 SBIGRU-HA、GRU、LSTM 在 108 号状态量的 F 得分分别为 0.9125，0.9339，0.9099，其中 SBIGRU-HA 与 GRU 的结果相差最大，达到 2.14%；SBIGRU-HA、GRU、LSTM 在 109 号状态量的 F 得分分别为 0.9124、0.9286、0.9328，其中 SBIGRU-HA 与 LSTM 的结果相差最大，达到 2.04%。图中 0、57 号状态量分别为平衡节点的幅值和相角，除去平衡节点，SBIGRU-HA 稳定性优于传统 GRU、LSTM，SBIGRU-HA 模型检测效果最差的点出现在 69 号，结果为 0.8363，而 GRU、LSTM 的结果分别为 0.8231、0.8007。换而言之，在 SBIGRU-HA 检测效果最差点的结果依旧好于 GRU、LSTM 的结果。而 LSTM 效果最差点出现在 58 号，其结果为 0.7412，SBIGRU-HA 检

测结果为 0.9269，在这一点 SBIGRU-HA 相对 LSTM 而言模型的检测 F 得分提升 18.57%，GRU 效果最差的点出现在 74 号，其结果为 0.7822，SBIGRU-HA 结果为 0.9014，SBIGRU-HA 在这一点的检测 F 得分比 GRU 高 11.92%。总体来说在综合指标 F 得分上 SBIGRU-HA 效果优于传统 GRU、LSTM，其能够相对准确地辨识虚假数据注入攻击的位置。

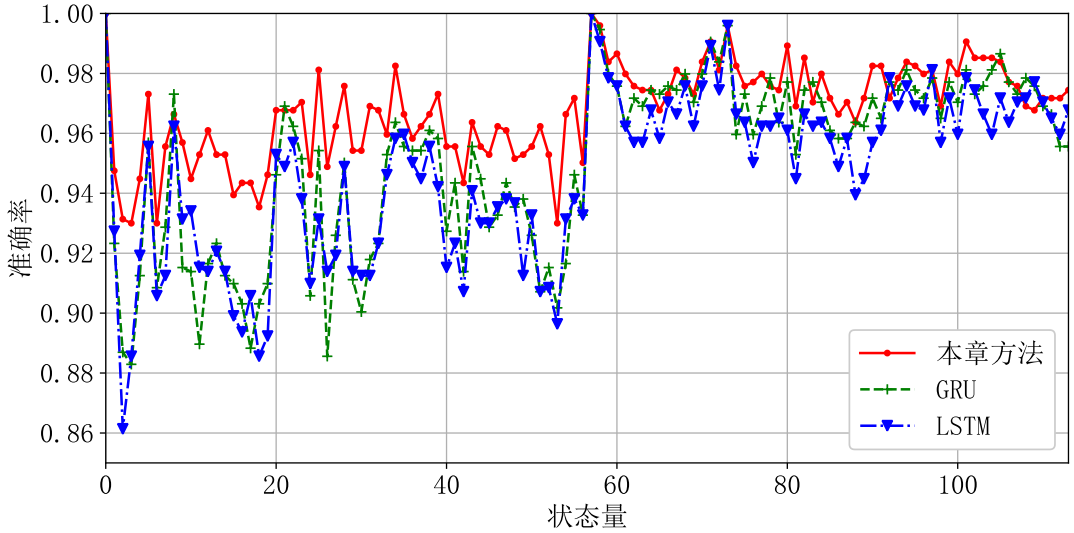


图 3-9 IEEE-57 节点系统各状态量检测的准确率

从图 3-9 中可以看出，LSTM 的检测准确率均在 86%以上，GRU 的检测准确率均在 88%以上，这表明其能够较好地区分出真实数据与虚假数据。而考虑时空特征并融合注意力的 SBIGRU-HA 模型检测准确率均在 92%以上，除去参考节点，检测效果最优的点都出现在 73 号，检测结果都达到 99%以上。LSTM、GRU、SBIGRU-HA 三者最高检测率和最低检测率相差分别达到 13%、11%、7%，很明显从稳定性而言，SBIGRU-HA 模型的检测效果优于 GRU、LSTM。但 SBIGRU-HA 在个别检测点的准确率效果却差于 LSTM、GRU，如 8、65、78、83、92、108、109 号，其中 LSTM 在 109 号状态量效果优于 SBIGRU-HA 最多，检测结果分别为 97.71%、96.77%，LSTM 在这一点也就比 SBIGRU-HA 高 0.94%；而 GRU 在 8 号状态量效果优于 SBIGRU-HA 最多，检测结果分别为 97.31%、96.64%，GRU 在这一点的检测结果也就比 SBIGRU-HA 高 0.67%，这说明 SBIGRU-HA 不仅在准确率上优于 GRU、LSTM，在稳定性上也更胜一筹。综上所述，SBIGRU-HA 能够准确地检测到系统中的虚假数据，有效保障了系统的安全与稳定。

同样，采用 3 层 BIGRU 进行检测，在 IEEE-57 节点完成仿真，SBIGRU-HA、GRU、LSTM 的 ROC 曲线如图 3-10 所示：

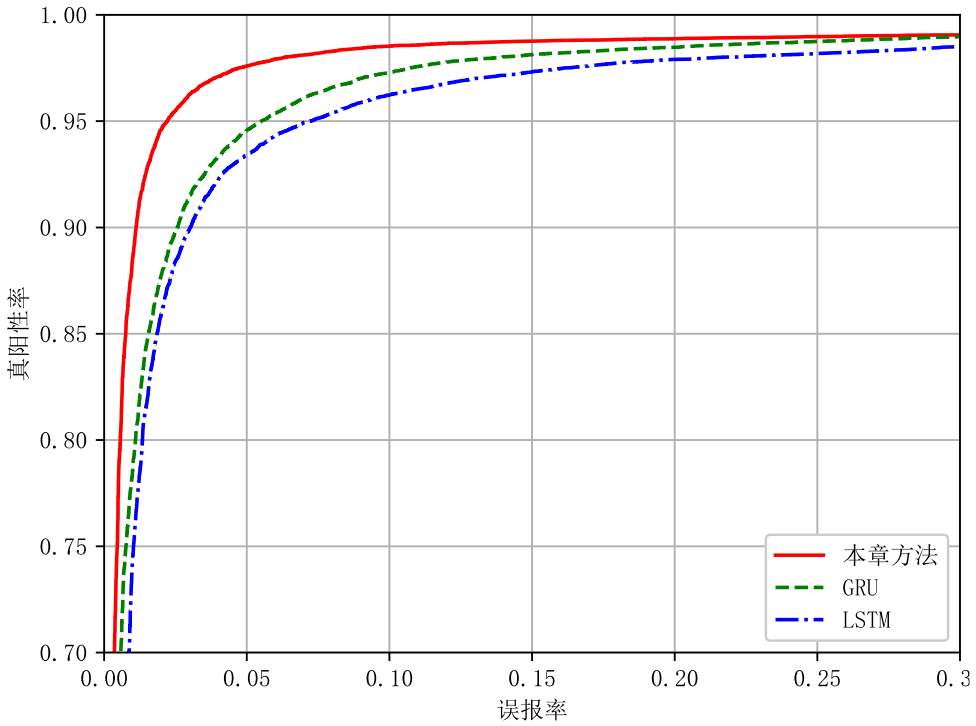


图 3-10 IEEE-57 节点系统 ROC 曲线

由图 3-10 可知，SBIGRU-HA 优于 GRU、LSTM。图中 SBIGRU-HA、GRU、LSTM 三种算法的 AUC 值分别为 0.9881、0.9841、0.9793。当误报率为 1%时，GRU 和 LSTM 真阳性率分别约为 0.9629、0.9373，SBIGRU-HA 的结果约为 0.9867，相较于 GRU 和 LSTM 分别提升 2.38%、4.88%，充分说明 SBIGRU-HA 模型的优越性。

3.3 本章小结

针对当前智能电网 FDIA 检测的时间数据相关性的问题，研究一种融合 SBIGRU 与三种注意力机制的检测模型，并在 IEEE-14 以及 IEEE-57 系统进行仿真实验。所采用数据集中的攻击模型不仅分别考虑已知全部拓扑信息、部分拓扑信息的情况，而且还考虑到 PMU 量测对攻击的影响。SBIGRU-HA 不仅从连续系统状态中呈现的时间数据相关性中学习，而且还考虑到数据的空间特征，在两个公共电网上的综合实验结果表明，SBIGRU-HA 在 IEEE-14 系统上相比于 GRU、LSTM 准确率分别提升 2.9%、2.74%，在 IEEE-57 系统则分别提升 1.67%、2.13%。在电网攻击检测中相比于误检，漏检对电网的危害更大，因此需更加关注综合指标 F 得分。SBIGRU-HA 在 IEEE-14 系统上相比于 GRU、LSTM 的 F 得分则分别提升 3.66%、3.68%，在 IEEE-57 系统则分别提升 2.52%、3.02%。综上，SBIGRU-HA

能更好完成 FDIA 检测。但在设计 SBIGRU-HA 模型时，未考虑训练时间、检测时间，未来将进一步权衡检测准确率、检测时间、模型复杂度和训练时间等多种因素，以及改进 BIGRU 的网络结构并采用更先进的注意力机制，降低模型的假阳性和阴性率。

第 4 章 虚假数据注入攻击下智能电网容侵控制

容侵是系统的最后一道防线，针对于无法有效排除的干扰，容侵可以在即使系统中某些组建遭到破坏后仍然可以为系统的整体提供完整或者降级的服务[73]。针对于电力系统中重在防护的现状，容侵技术的引入可以作为保护以及检测系统的有效补充。因此，本章对遭受攻击的系统数据完成数据恢复，从而维持系统稳定运行，亦可看作为一种容侵技术。针对电网物理系统的脆弱性，探究了一种基于 AE-WGANGP 的无监督状态恢复算法。考虑到数据的高维性和非线性相关特性，将自编码器应用于电力系统状态数据集的降维和特征提取。WGANGP 通过引入 Wasserstein 距离和梯度惩罚项以避免 GAN 在对抗训练中容易出现梯度消失和模式崩溃问题。在此基础上，我们将自动编码器引入到 GAN 框架下，该框架采用对抗训练，最终达到纳什平衡，通过捕获正常状态数据中的关键特征，学习数据间的空间关系，并完成状态数据的重构。实际系统中攻击数据稀少，且标记成本昂贵，该方法无需来自测量仪器的标记数据，只需无攻击的历史状态数据即可完成训练，因此是无监督学习。训练完成后，每当系统检测到 FDIA，只需将被攻击的状态数据剔除后送入 AE-WGANGP，即可得到恢复后的状态数据。

4.1 基于自编码器和带梯度惩罚的 Wasserstein 生成对抗网络模型

用于数据恢复的 AE-WGANGP 模型的具体结构如图 4-1 所示，其总体分为两部分，上半部分为由编码器和解码器组成的生成器，下半部分为判别器。

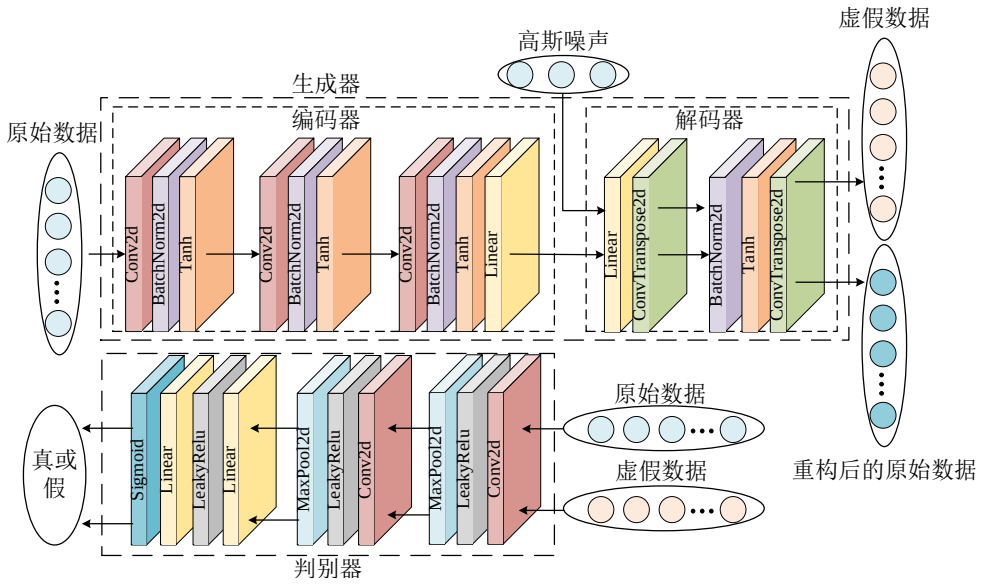


图 4-1 AE-WGANGP 模型

4.1.1 自编码器

近年来发展起来的深度学习方法可以有效地学习特征间的潜在模式，并将样本从原始特征空间映射到另一个易于识别的特征空间。自动编码器将输入特征编码到一个内隐层，然后通过最小化输入和输出数据之间的差异，将这一内隐层解码回原始输入特征。通过使用原始状态数据训练 AE，重建的输入将逐渐接近原始分布，其模型如图 4-2 所示。一般采用基于距离的函数（如欧氏距离、谱角距离）作为重构损失函数，通过反向传播进行网络训练。

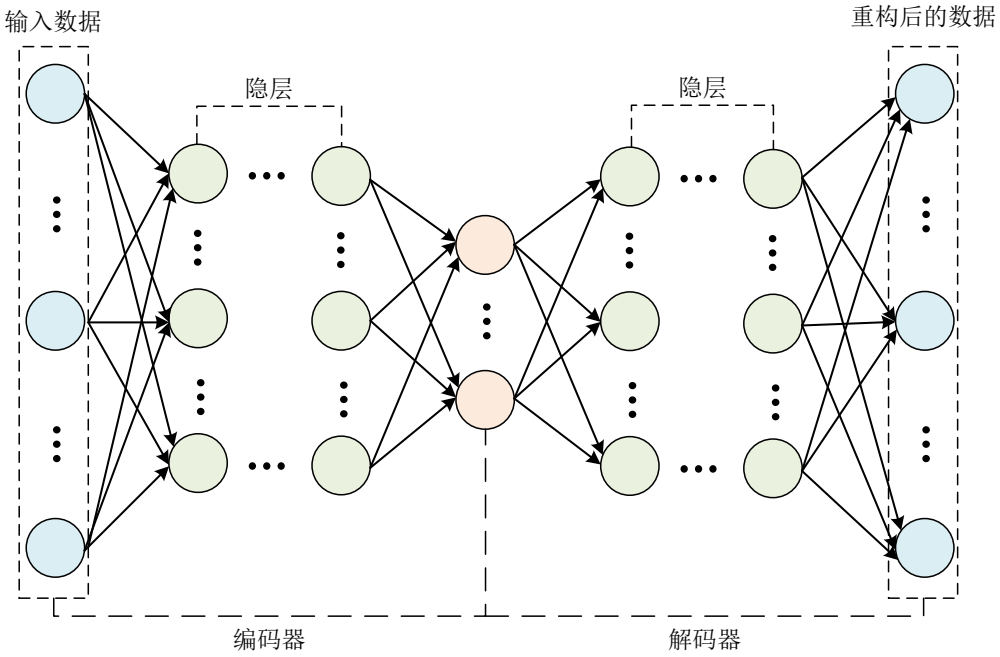


图 4-2 AE 模型

4.1.2 带梯度惩罚的 Wasserstein 生成对抗网络

GAN 的训练过程类似于一种博弈过程，其中生成器网络与判别器网络相互对抗。生成器网络 G 直接生成以随机噪声作为输入的样本，判别器网络 D 将生成的样本与真实样本区分开来。判别器的输出可以看作是其输入是真实样本而不是生成器网络得到的假样本的概率。

零和博弈可以更直观地描述生成器 G 和判别器 D 的训练过程，在网络训练过程中，生成器和判别器都会努力使自己的利益最大化，最终达到纳什平衡。这个过程驱动生成器尝试生成接近真实分布的样本，同时使判别器提高其确定样本真实性的能力。因此，生成器和判别器在训练过程中是同步进化的。当 GAN 收敛时，生成器生成的样本很难与真实样本区分。原始 GAN 中生成器与判别器的对抗关系定义如下：

$$\min_{G}\max_{D}V(G,D)=E_{x\sim p_{data}(x)}[\log D(x)]+E_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (4.1)$$

其中：$V(G,D)$为生成器和判别器的值函数，p_{data}为实际数据的分布，p_z是输入噪声的分布，x为真实数据，z为随机噪声。

然而，传统的 GAN 在实际应用中容易出现模型崩溃、训练困难等问题。这个问题主要是由于在 GAN 中使用 Jensen-Shannon(JS)散度来度量生成的数据分布与真实分布之间的距离。与 JS 散度相比，Wasserstein 距离可以平滑地反映两个不重叠或可忽略的重叠部分之间的距离，利用 Wasserstein 距离来度量生成的分布与真实数据分布之间的差异，可以在一定程度上解决训练不稳定的问题。Wasserstein 距离定义如下：

$$W(p_r,p_g)=\inf_{\gamma\sim\Pi(p_r,p_g)}E_{(a,b)\sim\gamma}(\|a-b\|) \quad (4.2)$$

其中：p_r为实际数据，p_g为生成数据，$W(p_r,p_g)$表示它们之间的 Wasserstein 距离，$\Pi(p_r,p_g)$是p_r和p_g分布组合起来的所有可能的联合分布的集合。对于每一个可能的联合分布γ，可以从中采样$(a,b)\sim\gamma$得到一个样本a和b，并计算出这对样本的距离$\|a-b\|$，同时计算该联合分布γ下样本对距离的期望值$E_{(a,b)\sim\gamma}(\|a-b\|)$，在所有联合分布中求期望值的最大下界$\inf_{\gamma\sim\Pi(p_r,p_g)}E_{(a,b)\sim\gamma}(\|a-b\|)$即为 Wasserstein 距离[74]。

然而，为了满足 Lipschitz 的连续条件，WGAN 将网络权值削减到一定范围内，间接限制了梯度信息。这样对于深度神经网络而言，就不能充分发挥其拟合能力。在判别器的损失函数中加入梯度惩罚项，WGAN-GP 模型生成器和判别器的损失函数为：

$$L_g=-E_{\tilde{x}\sim p_g}[f_w(\tilde{x})] \quad (4.3)$$

$$L_d=E_{\tilde{x}\sim p_g}[f_w(\tilde{x})]-E_{x\sim p_r}[f_w(x)]+\lambda E_{\hat{x}\sim p_{\hat{x}}}[(\|\nabla_{\hat{x}}f_w(\hat{x})\|_2-1)^2] \quad (4.4)$$

$$\hat{x}=\varepsilon x+(1-\varepsilon)\tilde{x} \quad (4.5)$$

其中：$f_w(\cdot)$为 WGAN-GP 模型中判别器的可微函数，λ为正则化系数，$\|\cdot\|_2$为 2 范数，ε为 0~1 之间随机数，∇为梯度，$\hat{x}$为实际数据与生成数据之间的随机插值结果。损失函数加入梯度惩罚项后，判别器能判断输入数据的真实性，还可辨别类别标签与输入数据的真实类别是否匹配。

4.1.3 生成器

所采用的 AE 模型分为两部分：编码器和解码器。编码器结构如下[Conv2d-BatchNorm2d-Activation(tanh)-Conv2d-BatchNorm2d-Activation(tanh)-Conv2d-BatchNorm2d-Activation(tanh)-Linear]，解码器结构如下[Linear-ConvTranspose2d-BatchNorm2d-Activation(tanh)-ConvTranspose2d]。

Conv2d、ConvTranspose2d 分别为二维卷积和二维反卷积层，Conv2d 的卷积核大小设置为 3，步幅为 2，填充为 1，ConvTranspose2d 的卷积核大小设置为 1，步幅以及填充为默认值，BatchNorm2d 为批量标准化层，Activation(tanh)为非线性激活函数 tanh，Linear 表示线性层。

4.1.4 判别器

判别器结构如下 [Conv2d-Activation(LeakyRelu)-MaxPool2d-Conv2d-Activation(LeakyRelu)-MaxPool2d-Linear-Activation(LeakyRelu)-Linear-Activation(Sigmoid)]。Conv2d 卷积核、步幅和填充都为 1，MaxPool2d 表示最大池化操作，池化窗口为 2×2，Activation(LeakyRelu)为非线性激活函数 LeakyRelu，其斜率设置为 0.2，Activation(Sigmoid)为 Sigmoid 激活函数，用于将输入映射到[0,1]之间。

4.1.5 对抗训练

AE-WGANGP 的训练过程主要是生成器和判别器之间的对抗训练过程。一方面，判别器试图拟合真实样本和生成样本分布之间的 Wasserstein 距离。另一方面，生成器试图最小化 Wasserstein 距离，使真实样本和生成样本的分布更接近。此外，生成器减少了真实样本和生成样本之间的平均绝对误差，从而使它们更接近。

AE-WGANGP 对抗训练的流程图如图 4-3 所示，其步骤如下：

1）将原始数据输入判别器 D 得到其输出记为 real_out，生成随机高斯噪声并将其输入解码器，得到其生成的假数据 fake_data，再将 fake_data 送入判别器 D 得到输出记录为 fake_out。

2）按公式 4.5 对原始数据和 fake_data 进行随机插值得到 x_hat，x_hat 输入判别器 D 得到 pre_hat，其后按公式 4.4 中的第三项计算梯度惩罚项 gradient_penalty，取 fake_out 的平均值减 real_out 的平均值加梯度惩罚项 gradient_penalty 即得判别器损失，而后反向传播，参数更新。

3）将原始数据输入判别器 D 得到其输出记为 real_out，生成随机高斯噪声

并将其输入解码器，得到其生成的假数据 fake_data，再将 fake_data 送入判别器 D 得到输出记录为 fake_out。

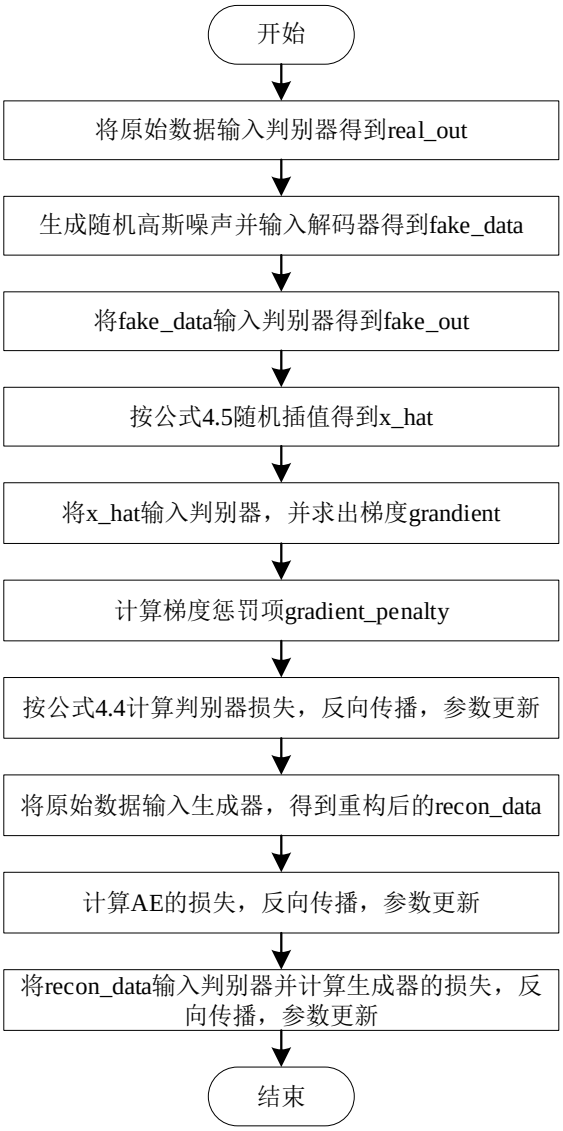


图 4-3 AE-WGANGP 训练流程图

4）按公式 4.5 对原始数据和 fake_data 进行随机插值得到 x_hat，x_hat 输入判别器 D 得到 pre_hat，其后按公式 4.4 中的第三项计算梯度惩罚项 gradient_penalty，取 fake_out 的平均值减 real_out 的平均值加梯度惩罚项 gradient_penalty 即得判别器损失，而后反向传播，参数更新。

5）将原始数据输入生成器 G，得到重构后的输入 recon_data，计算原始数据和 recon_data 的均方差即为 AE 的损失，反向传播，参数更新。

6）根据公式 4.3 求出生成器的损失值。具体而言，将生成器的重构输出 recon_data 输入判别器 D，并取其输出负值的平均值即为损失，然后再反向传播，

梯度更新即可完成一次训练。

对抗训练的部分代码如下：

AE-WGANGP 对抗训练部分代码

```
real_out=D(data)
z=torch.randn(batch_size,32,1,1)
fake_data=vae.decoder_fc(z)
fake_data=vae.decoder(fake_data)
fake_out=D(fake_data)
alpha=torch.rand((batch_size,1,1,1))
x_hat=alpha * data + (1 - alpha) * fake_data
pred_hat=D(x_hat)
gradients=torch.autograd.grad(outputs=pred_hat,inputs=x_hat,
                              grad_outputs=torch.ones(pred_hat.size()),
                              create_graph=True,retain_graph=True,
                              only_inputs=True)[0]
gradient_penalty=lambda_*((gradients.view(gradients.size()[0],-1).norm(2,1)-1)
                          **2).mean()
d_loss=torch.mean(fake_out)-torch.mean(real_out)+gradient_penalty
d_loss.backward()
optimizerD.step()
recon_data=vae(data)
vae.zero_grad()
vae_loss=loss_function(recon_data, data)
vae_loss.backward()
optimizerVAE.step()
vae.zero_grad()
output=D(recon_data)
errVAE=torch.mean(-output)
errVAE.backward()
D_G_z2=output.mean().item()
optimizerVAE.step()
```

4.2 实验验证及结果分析

4.2.1 数据集构建

数据集构建采用交流电力系统模型，相比较直流电力系统模型考虑了无功潮流和节点无功以及并联充电电纳。在智能电网中，状态估计的处理对象是一个时间段上的高维系统，一般采用最小二乘法。系统的非线性测量方程可表示为

$$z = h(x) + e \tag{4.6}$$

其中：$z \in \mathbb{R}^{m\times 1}$ 为测量向量；$h(\cdot)$ 为非线性测量方程；$x \in \mathbb{R}^{n\times 1}$ 为母线电压的幅值和

相位角组成的状态变量；e 是测量误差；$h(x)$ 可以表示为

$$h(x)=\begin{bmatrix} P_{ij} \\ Q_{ij} \\ P_i \\ Q_i \\ U_i \end{bmatrix}, i=1,...,N; \forall j \in \Omega_i \tag{4.7}$$

其中：P_{ij} 为母线 i 到母线 j 的有功功率流，Q_{ij} 为母线 i 到母线 j 的无功功率流，P_i 为母线 i 的有功功率注入，Q_i 为母线 i 的无功功率注入，U_i 为母线 i 的电压幅值，Ω_i 为与母线 i 相连的所有母线的集合。

状态估计的解可以通过加权最小二乘法得到：

$$\hat{x}=\arg\min[z-h(x)]^T R^{-1}[z-h(x)] \tag{4.8}$$

其中：$R=diag(\sigma_1^2,...,\sigma_m^2)\in\mathbb{R}^{m\times m}$ 为测量误差向量的协方差；σ_i^2 为第 i 个仪表的测量误差的方差；$\hat{x}\in\mathbb{R}^{n\times 1}$ 为状态估计向量。

公式 4.8 可以通过高斯-牛顿算法迭代求解：

$$x^{l+1}=x^l+[G(x^l)]^{-1}\{H^T(x^l)R^{-1}[z-h(x^l)]\} \tag{4.9}$$

$$G(x^l)=H^T(x^l)R^{-1}H(x^l) \tag{4.10}$$

其中：l 为迭代索引；x^l 是迭代 l 次时的解向量;$H\in\mathbb{R}^{m\times n}$ 是雅可比矩阵。加权最小二乘状态估计在电力系统中应用广泛，但也存在一定的局限性。因此，在实践中，状态估计后还需要进行不良数据检测。

传统的不良数据检测方法，如 LNR 检验和卡方检验，都是基于加权最小二乘状态估计的结果进行的。利用 SCADA 系统提供的冗余信息对不良数据进行判断和处理，残差的二范数如下：

$$\begin{aligned} r&=\left\|z-h(\hat{x})\right\|_2 \\ &=\left\|h(x)-h(\hat{x})+e\right\|_2 \end{aligned} \tag{4.11}$$

然后将所得残差与阈值 τ 进行比较，如果 $r>\tau$，则检测结果为存在攻击；反之，测量结果则是正常的。

FDIA 一般构建攻击向量 a 为：

$$\begin{aligned} a&=h(\hat{x}+c)-h(\hat{x}) \\ &=h(c) \end{aligned} \tag{4.12}$$

其中：c表示攻击者注入的状态偏差。攻击者发起 FDIA 后，控制中心接收到的篡改测量数据z_a为

$$\begin{aligned} z_a &= z + a \\ &= z + h(\hat{x} + c) - h(\hat{x}) \end{aligned} \tag{4.13}$$

相应的残差为

$$\begin{aligned} r_a &= \left\| z_a - h(\hat{x}_a) \right\|_2 \\ &= \left\| z + h(\hat{x} + c) - h(\hat{x}) - h(\hat{x}_a) \right\|_2 \\ &= \left\| z - h(\hat{x}) \right\|_2 \\ &= r \end{aligned} \tag{4.14}$$

其中：$\hat{x}_a = \hat{x} + c$表示攻击后的状态数据。由公式 4.14 可知，FDIA 后的残差r_a与未受攻击时的残差r相同，因此 FDIA 是隐形的，不会触发 BDD 机制。

由于目前并没有公开的 FDIA 数据集以及相应的攻击前的原始数据集，数据集在 MATPOWER 工具包下模拟生成。攻击分为两种，一种为随机单节点攻击，另一种为随机多节点攻击。生成两个数据集，第一个数据集中为随机单节点攻击，数据集中为攻击前以及攻击后的状态数据各 2200 条。第二个数据集中为随机多节点攻击，其中也包含攻击前以及相应的攻击后的状态数据分别有 2200 条。生成的状态量都需叠加均值为 0，标准差为 0.005 的高斯噪声，从而模拟真实情况。IEEE-14 状态数据生成的流程图如图 4-4 所示，其步骤如下：

1）生成随机有功、无功负荷。有功无功负荷在其基准值的 0.9-1.1 之间随机波动。使用 MATLAB 中的 rand 函数，假设有功、无功的其基准值分别为 mean_p、mean_q，则随机生成有功的表达式为：$(0.2 \times rand(n,1) + 0.9 \times ones(n,1)) \times mean\_p$，生成无功的表达式为：$(0.2 \times rand(n,1) + 0.9 \times ones(n,1)) \times mean\_q$。

2）潮流计算得到状态值以及量测值z。

3）生成随机的攻击节点 attack_bus。由于电力系统的特性，若想完成 FDIA 则需避开参考节点，采用 MATLAB 中的 randperm 函数生成随机攻击节点，若是单节点攻击则 attack_bus=randperm(13,1)+1；三节点攻击则 attack_bus=randperm(13,3)+1。其中 randperm 函数的第一个参数 13 表示其生成的数据范围为[1,13]，第二个参数表示生成数据的个数，比如 randperm(13,1)表示生成一个随机数其范围在[1,13]。

4）生成随机攻击偏差。使用 MATLAB 中的 normrnd 函数。若是单节点攻击

设置其为 normrnd(0,0.1,[1,2])；三节点攻击设置其为 normrnd(0,0.1,[1,6])。其中 normrnd 函数的第一个参数表示均值，第二个参数表示标准差，第三个参数为生成数据的形状。normrnd(0,0.1,[1,2])就表示生成一个行向量，形状为1×2，其满足均值为 0，标准差为 0.1 的高斯分布；normrnd(0,0.1,[1,6])就表示生成一个行向量，形状为1×6，其满足均值为 0，标准差为 0.1 的高斯分布。

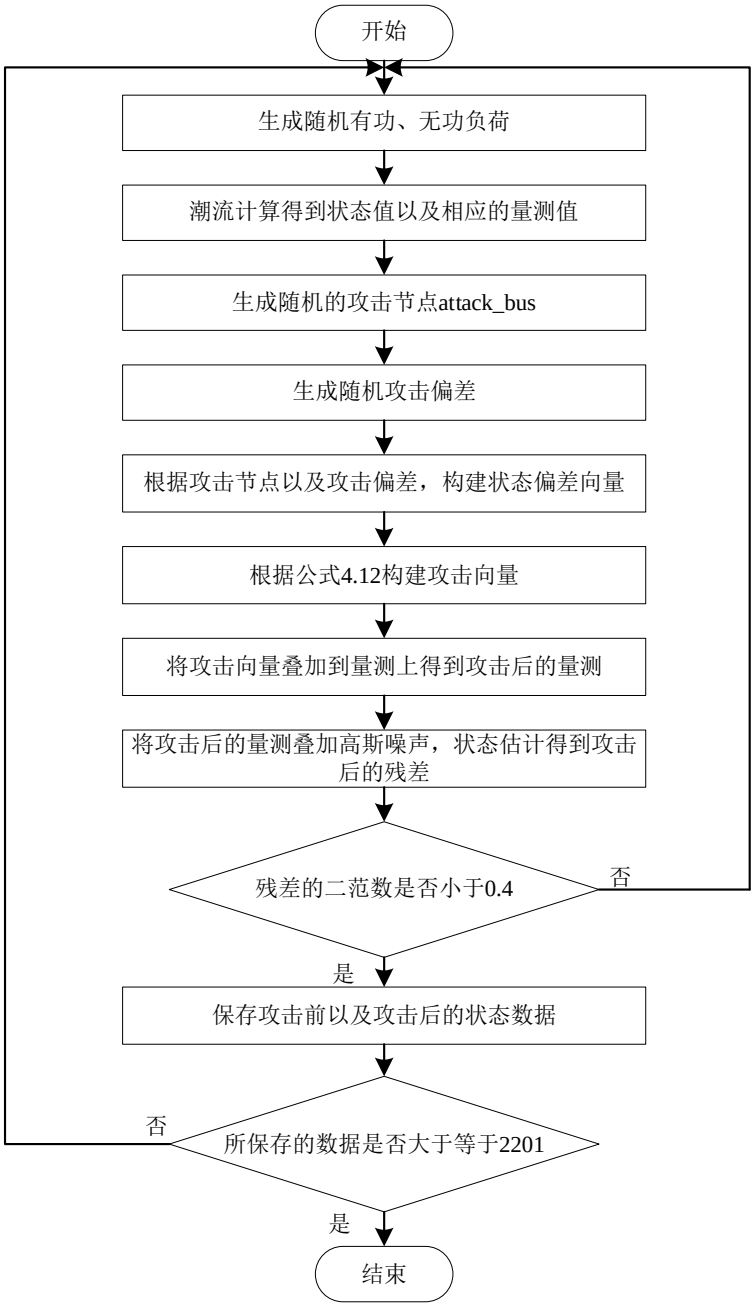


图 4-4 数据生成流程图

5）根据攻击节点以及攻击偏差，构建状态偏差向量 c。以单节点攻击为例，运行 normrnd(0,0.1,[1,2])，得到攻击偏差 [0.0707，-0.0585]，运行

attack_bus=randperm(13,1)+1，得到攻击节点 5，则此时的状态偏差向量 $c=[0,0,0,0,0.0707,0,0,0,0,0,0,0,0,0,0,0,-0.0585,0,0,0,0,0]^T$。

6）根据公式 4.12 构建攻击向量 a，由公式 4.13 得到攻击后的量测 z_a，将 z_a 叠加均值为 0，标准差为 0.001 的高斯噪声，然后状态估计得到攻击后的残差 r_a。

7）判断攻击后残差 r_a 的二范数是否小于 0.4。若是，则保存攻击前以及攻击后的状态数据；反之，则回到步骤 1）。实际上，在此系统中量测值数量 $m=119$，状态值数量 $n=28$，若此时显著性水平设置为 0.05，冗余度 $k=m-n$，查阅卡方分布表可得此时阈值为 113.14，远远大于我们设置的阈值 0.4。

8）判断所保存的状态数据数量是否大于等于 2201。若是，则结束循环；反之，则回到步骤 1）。

4.2.2 攻击结果分析

1）单节点攻击

在 IEEE-14 节点系统上，使负荷在其基准值的 0.9-1.1 之间波动，潮流计算，得到状态值以及量测值，进行随机单节点攻击，获得攻击后的状态值以及量测值，而后分别用攻击前的量测 z 和攻击后的量测 z_a 分别进行状态估计得到攻击前的残差以及攻击后的残差。数据生成的残差如图 4-5 所示：

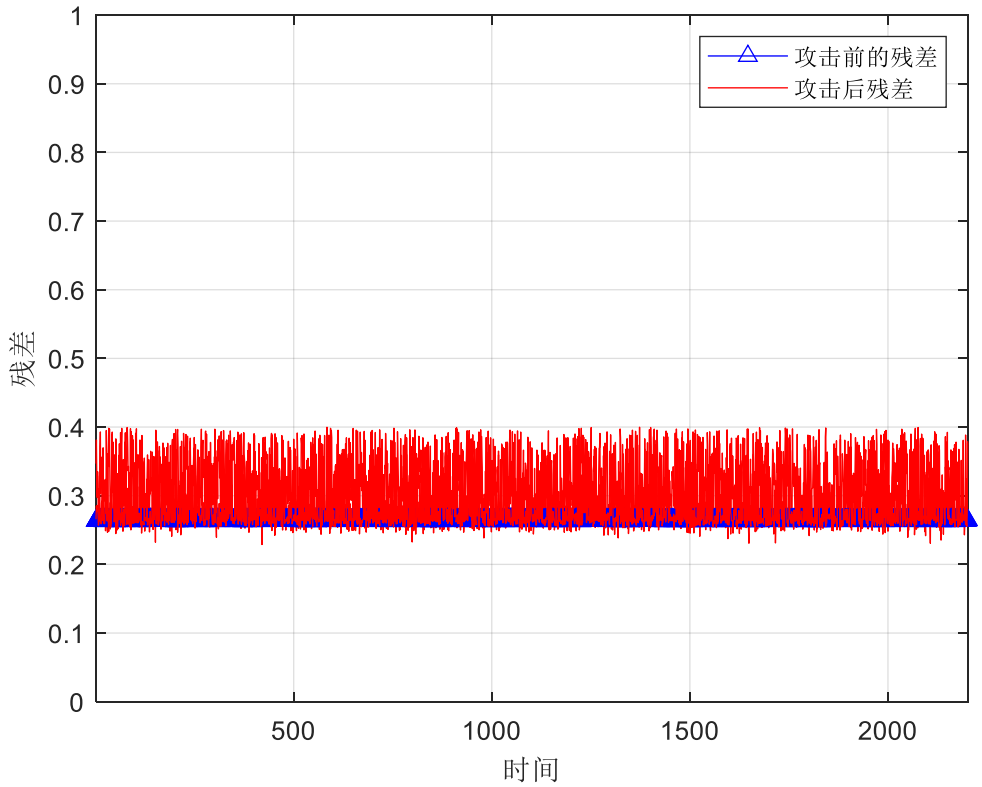


图 4-5 单节点攻击前、后残差

由图 4-5 可知，攻击后的系统残差均小于设置阈值 0.4，因此能成功完成 FDIA。

随后进行一次随机单节点攻击，并对比攻击前后系统的状态。运行 normrnd (0,0.1,[1,2])，得到攻击偏差[0.0707，-0.0585]，运行 attack _ bus=randperm(13,1)+1，得到攻击节点 5，则此时的状态偏差向量 $c=[0,0,0,0,0.0707,0,0,0,0,0,0,0,0,0,0,0,0,-0.0585,0,0,0,0,0]^T$，潮流计算得到状态真值为[1.06,1.043,1.0192,1.0154,1.0196,1.0144,1.0185,1.0583,0.99648,0.99223,0.99987,0.99887,0.99313,0.97499,0,-4.0862,-10.03,-8.6465,-7.4311,-13.014,-10.964,-9.8264,-12.962,-13.271,-13.269,-13.896,-13.964,-14.541]，攻击后的状态值为[1.06,1.043,1.0192,1.0154,1.0903,1.0144,1.0185,1.0583,0.99648,0.99223,0.99987,0.99887,0.99313,0.97499,0,-4.0862,-10.03,-8.6465,-10.785,-13.014,-10.964,-9.8264,-12.962,-13.271,-13.269,-13.896,-13.964,-14.541]，根据公式 4.12、4.13 构建攻击向量 a 得到攻击后的量测 z_a，将 z 和 z_a 叠加均值为 0，标准差为 0.001 的高斯噪声，然后分别用攻击前以及攻击后的量测进行状态估计，结果如图 4-6 所示，节点 5 的电压幅值由攻击前的 0.98p.u 上升到 1.05p.u，电压相角由原先的-7.98° 下降到-11.23° 。

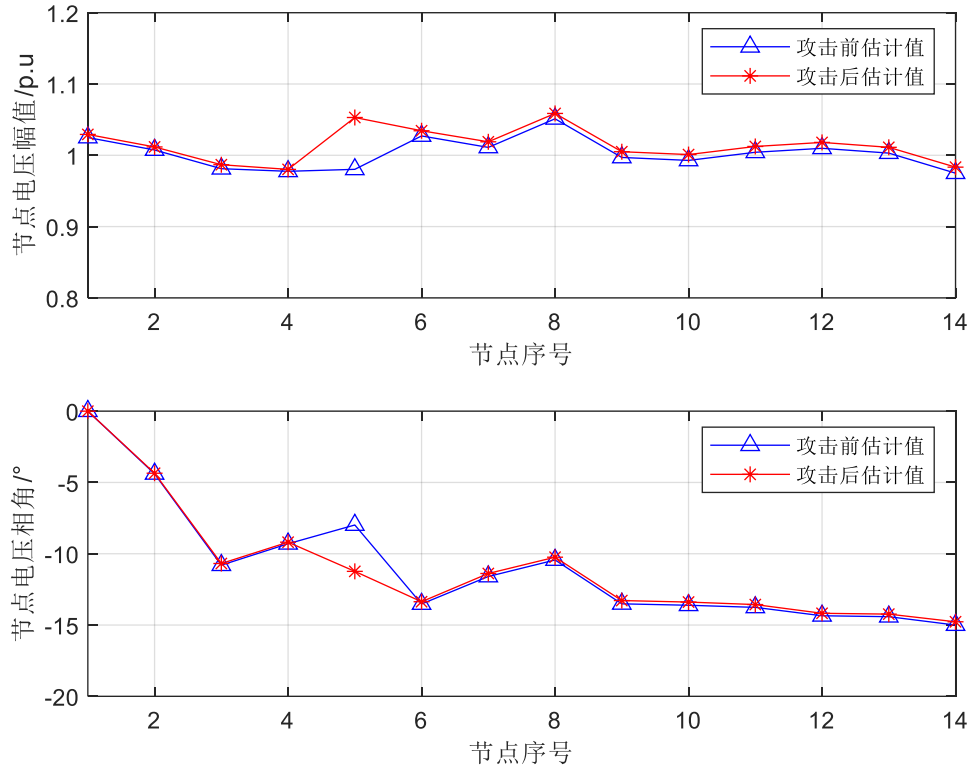


图 4-6 单节点攻击效果图

2）多节点攻击

数据生成的残差如图 4-7 所示，由图可知，攻击前的残差依旧小于设置阈值

0.4，因此能成功完成 FDIA。

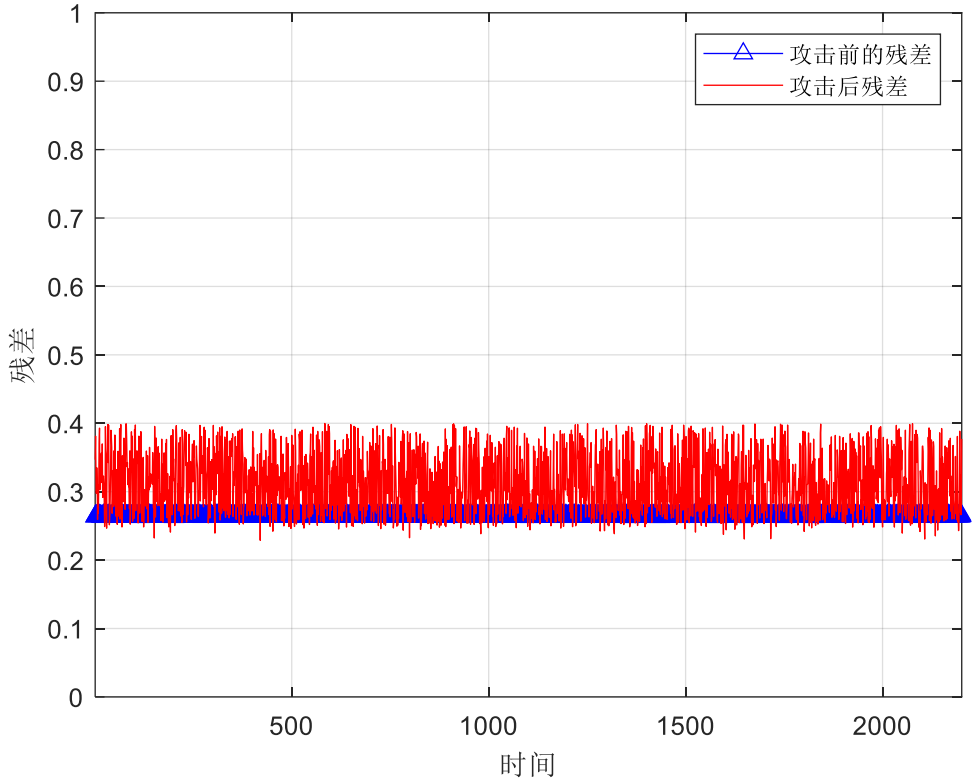


图 4-7 多节点攻击前、后残差

同样采取多节点攻击，查看系统是否发生状态偏移。运行 normrnd(0,0.1,[1,6])，得到攻击偏差[0.0511，-0.04634，-0.10257,0.0296，0.1071，-0.016]，运行 attack _ bus=randperm(13,3)+1，得到攻击节点[2,3,14]，则此时的状态偏差向量 $c=[0,0.0511,-0.04634,0,0,0,0,0,0,0,0,0,0,-0.10257,0,0.0296,0.1071,0,0,0,0,0,0,-0.016]^T$，电力系统潮流计算得到状态真值为[1.06,1.0431,1.0179,1.0156,1.0198,1.0177,1.0205,1.06,0.99956,0.99526,1.003,1.0023,0.9969,0.9791,0,-4.0475,-10.043,-8.6915,-7.479,-13.231,-11.366,-10.564,-13.327,-13.663,-13.575,-14.107,-14.21,-14.948]，攻击后的状态值为[1.06,1.0942,0.97156,1.0156,1.0198,1.0177,1.0205,1.06,0.99956,0.99526,1.003,1.0023,0.9969,0.87653,0,-2.3522,-3.9051,-8.6915,-7.479,-13.231,-11.366,-10.564,-13.327,-13.663,-13.575,-14.107,-14.21,-15.867]，根据公式 4.12、4.13 得到攻击向量 a 以及攻击后的量测 z_a，叠加均值为 0，标准差为 0.001 的高斯噪声后分别进行电力系统状态估计，其结果如图 4-8 所示，节点 2 的电压幅值由攻击前的 1.00p.u 上升至 1.06p.u，电压相角由原先的-4.3503° 上升至-2.5756° 。节点 3 的电压幅值由攻击前的 0.97p.u 下降至 0.93p.u，电压相角由原先的-10.8099° 上升至-4.0381° 。节点 14 的电压幅值由攻击前的 0.97p.u 下降至 0.87p.u，电压

相角由原先的-15.4048°下降至-16.1983°。

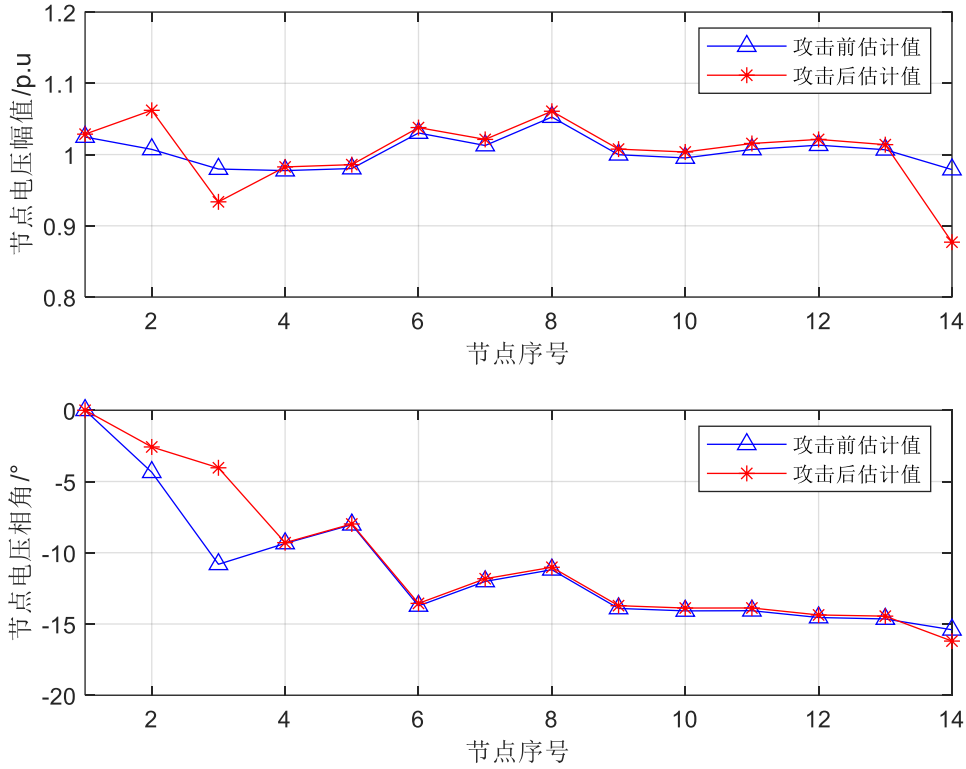


图 4-8 多节点攻击效果图

4.2.3 评价指标及实验设置

采用回归任务中最常用的评价指标：均方根误差(Root Mean Square Error，RMSE)、平均绝对误差(Mean Absolute Error，MAE)和平均绝对百分比误差(Mean Absolute Percentage Error，MAPE)，RMSE、MAE 和 MAPE 的值越小，意味着预测值与实际值之间的差距也就越小，RMSE、MAE、MAPE 的数学表达式如下所示：

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{t=1}^{n}(x_r(t)-x_p(t))^2}{n}} \tag{4.15}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{t=1}^{n}\left|x_r(t)-x_p(t)\right| \tag{4.16}$$

$$\mathrm{MAPE}=\frac{1}{n}\sum_{t=1}^{n}\frac{\left|x_r(t)-x_p(t)\right|}{x_r(t)}\times 100\% \tag{4.17}$$

其中：n 表示样本数，$x_r(t)$ 表示 t 时刻状态值的真实值，$x_p(t)$ 表示 t 时刻状态值的预测值。

IEEE-14 系统中共有状态量 28 个，其中电压幅值和相角各 14 个，在使用无攻击的状态量训练 AE-WGANGP 之前，首先对状态数据进行 z-score 标准化，公

式如下：

$$\hat{x}_j^{(i)} = \frac{x_j^{(i)} - \mu_j}{\upsilon_j} \tag{4.18}$$

$$\mu_j = \frac{1}{n}\sum_{i=0}^{n-1} x_j^{(i)} \tag{4.19}$$

$$\upsilon_j = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(x_j^{(i)} - \mu_j)^2} \tag{4.20}$$

其中：$x_j^{(i)}$ 表示标准化之前的状态量，$\hat{x}_j^{(i)}$ 表示标准化之后的状态量，i, j 分别表示行和列，μ_j 表示第 j 列的均值，υ_j 表示第 j 列的标准差，相应的模型最终的预测值要进行逆标准化从而得到原始分布的数据。实验平台环境为 Windows 11-X64 操作系统，显卡为 NVIDIA-GeForce-GTX-3060，16GB 运行内存，处理器为 12th Gen Intel(R) Core(TM) i5-12600KF 3.70 GHz。在 Pytorch1.2.0 框架、CUDA11.1 环境下搭建模型，评价指标采用 sklearn 库中现有函数完成，实验中 batchsize 为 128，epoch 为 200，AE-WGANGP 模型中的正则化系数 λ 设置为 10，优化器均使用 Adam，学习率均为 0.001，为进一步说明模型优越性，选取 AE、AE-GAN 作为对比试验，其中 AE-WGANGP、AE-GAN 中生成器的网络结构与 AE 中相同，AE-WGANGP 和 AE-GAN 中判别器的网络结构也相同，所有实验均在同一工况下进行。完成无监督状态量重构后，对于检测到的 FDIA，只需将被攻击的状态数据剔除，而后送入完成训练的 AE-WGANGP 模型即可得到恢复后的状态数据。

4.2.4 实验结果与分析

实验所使用的数据集包含攻击前的正常状态量 2200 条，以及攻击后受损的状态量 2200 条。使用前 2000 条无攻击的数据给 AE-WGANGP 进行对抗训练，完成无监督状态数据重构，将剩余的 200 条受损的状态数据剔除被攻击的状态量，即将这些受损的状态数据设为 0，而后将状态数据送入训练好的 AE-WGANGP 完成数据恢复。

1）单节点攻击

在时刻 1 采用随机单节点攻击，将受损的电压幅值设为零后，送入 AE、AE-GAN、AE-WGANGP 完成电压的恢复，相应的电压幅值恢复结果分别如图 4-9、4-10、4-11 所示。

由图 4-9、4-10、4-11 可知，异常点为 9 号节点，其值约为 0.89，三个图中

真实值略微不同是因为添加了均值为 0，标准差为 0.005 的高斯噪声以模拟测量误差。总体而言，AE-WGANGP 和 AE 拟合效果较好，AE-GAN 效果相对较差。在受损节点，AE-WGANGP、AE-GAN、AE 的恢复值都比较接近真实值，使系统稳定运行。

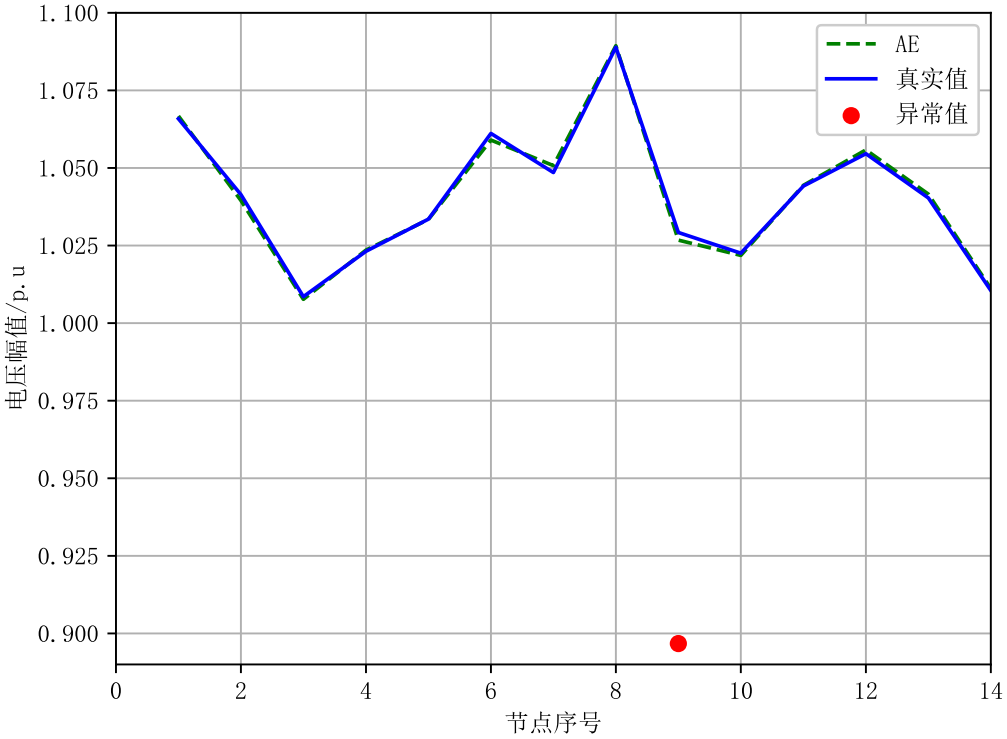


图 4-9 1 时刻单节点攻击后 AE 电压幅值恢复图

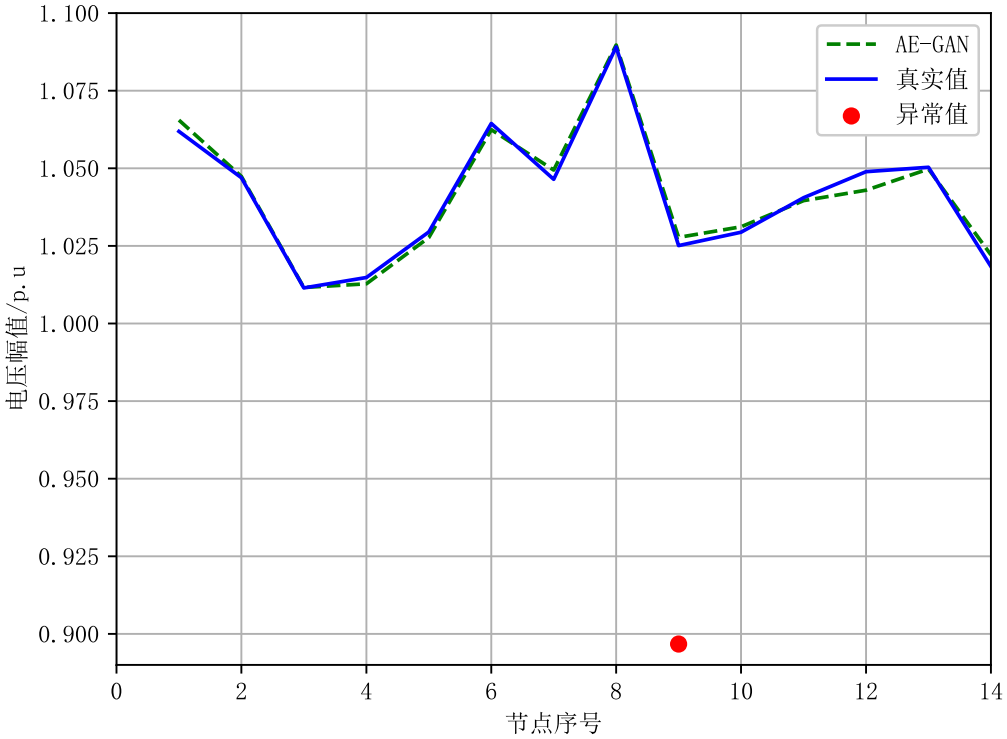


图 4-10 1 时刻单节点攻击后 AE-GAN 电压幅值恢复图

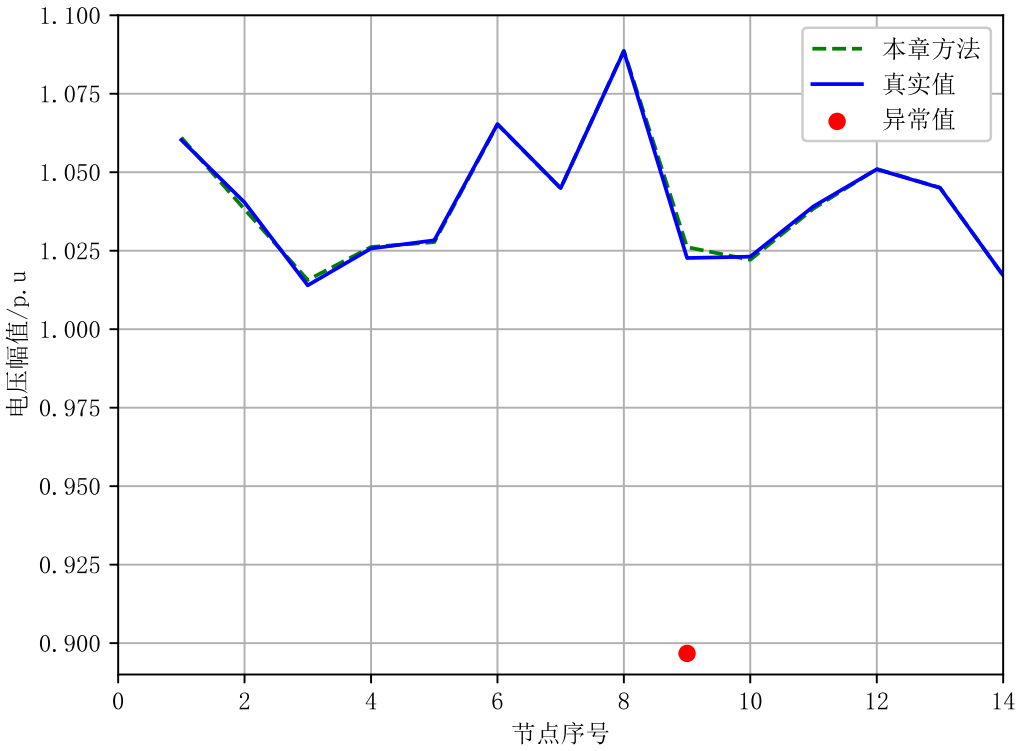


图 4-11 1 时刻单节点攻击后 AE-WGANGP 电压幅值恢复图

同样，时刻 1 电压相角恢复结果如图 4-12 所示：

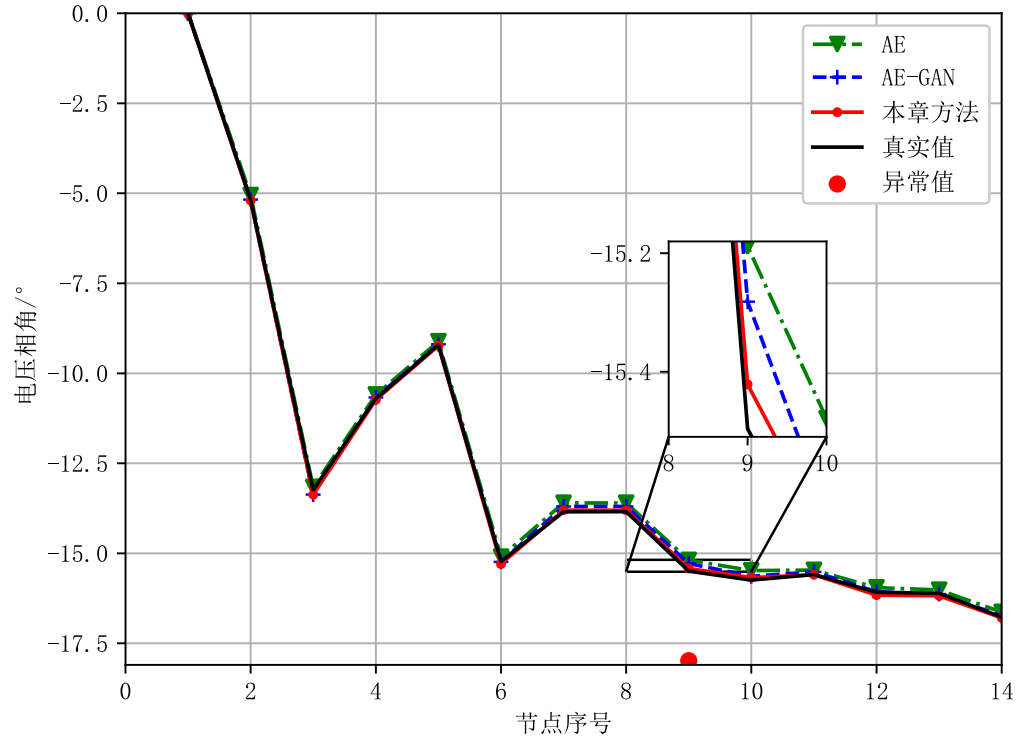


图 4-12 1 时刻单节点攻击后电压相角恢复图

由图可知，9 号节点电压相角异常，单看这一点的拟合情况，AE-WGANGP 拟合效果最好，AE-GAN 次之，AE 拟合效果较差。而在其他节点，三种算法均

能较好的拟合真实值。

由于 1 号节点为参考节点，无法被攻击，因此以 2 号节点为例展示其电压幅值在 200 个时刻中遭受随机单节点攻击后的电压幅值情况以及相应的恢复情况，AE、AE-GAN、AE-WGANGP 结果分别如图 4-13、4-14、4-15 所示：

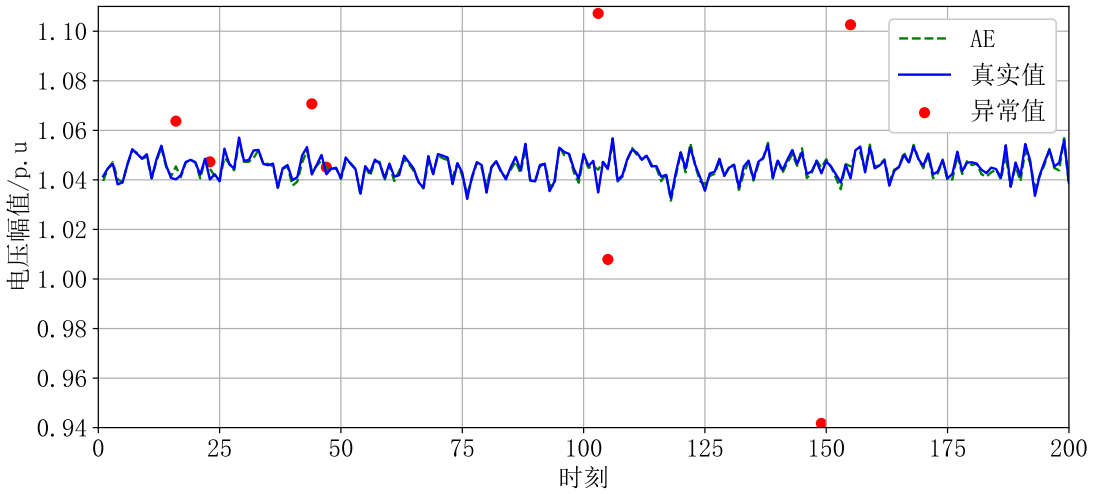


图 4-13 单节点攻击后 AE 电压幅值恢复图

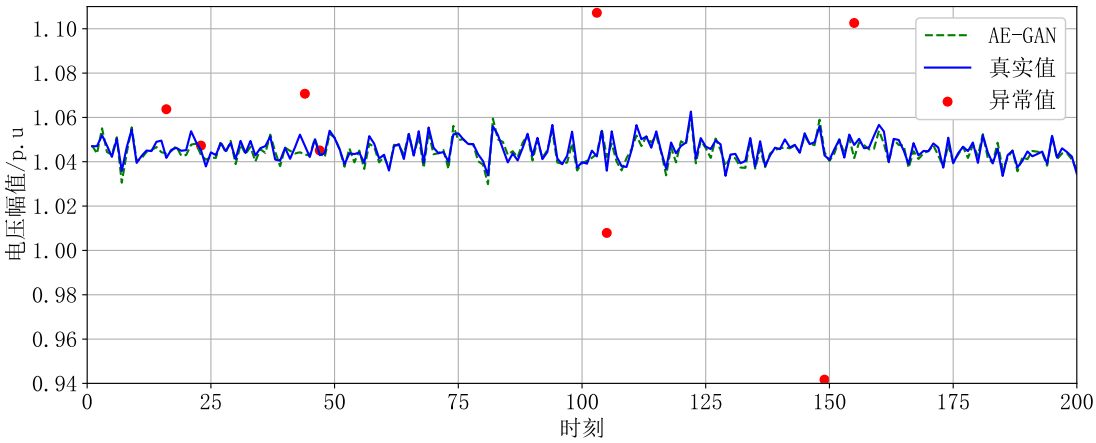


图 4-14 单节点攻击后 AE-GAN 电压幅值恢复图

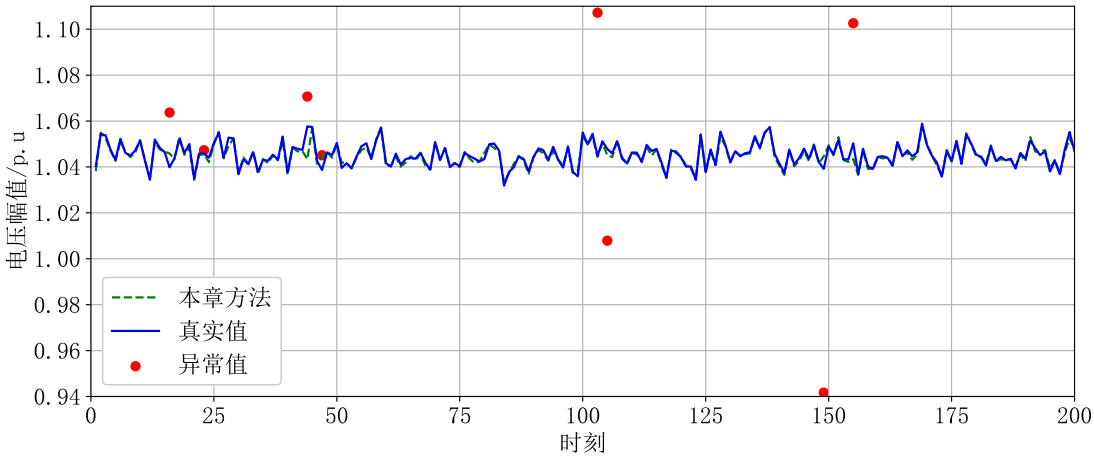


图 4-15 单节点攻击后 AE-WGANGP 电压幅值恢复图

图 4-13、4-14、4-15 中节点二分别在 16、23、44、47、103、105、149、155 时刻遭受攻击，其异常值分别为 1.0637、1.0473、1.0707、1.0451、1.1072、1.0079、0.94171、1.10260。攻击注入后，系统的电压幅值明显偏离正常值，经过 AE-WGANGP 修正后，与攻击前的值基本拟合。

2 号节点其电压相角在 200 个时刻中所受攻击以及 AE、AE-GAN、AE-WGANGP 模型恢复效果分别如图 4-16、4-17、4-18 所示：

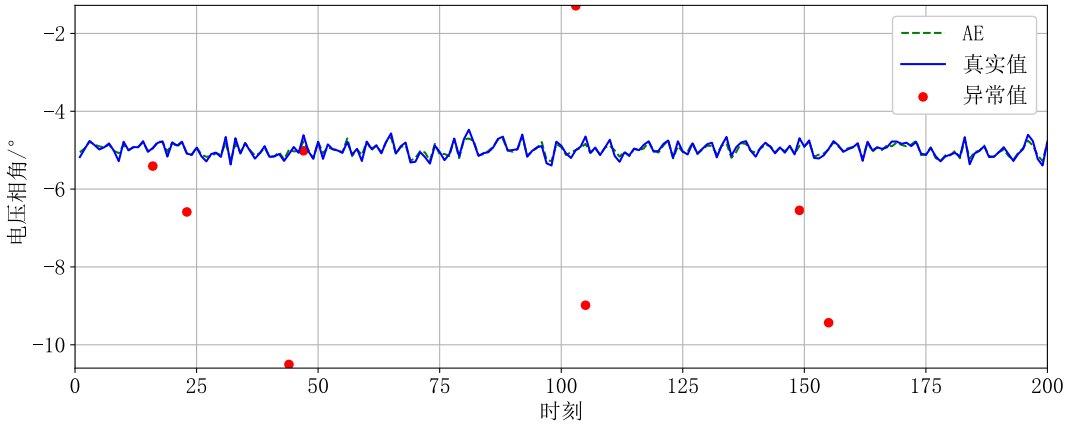


图 4-16 单节点攻击后 AE 电压相角恢复图

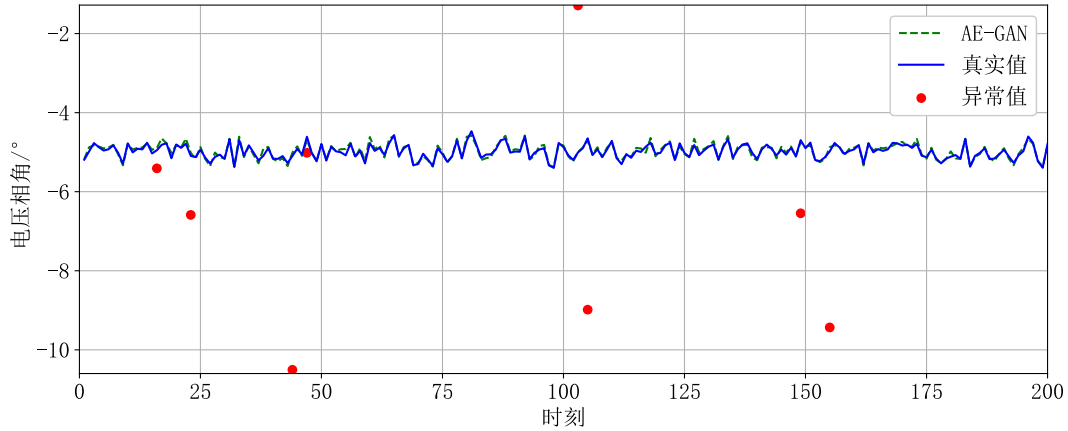


图 4-17 单节点攻击后 AE-GAN 电压相角恢复图

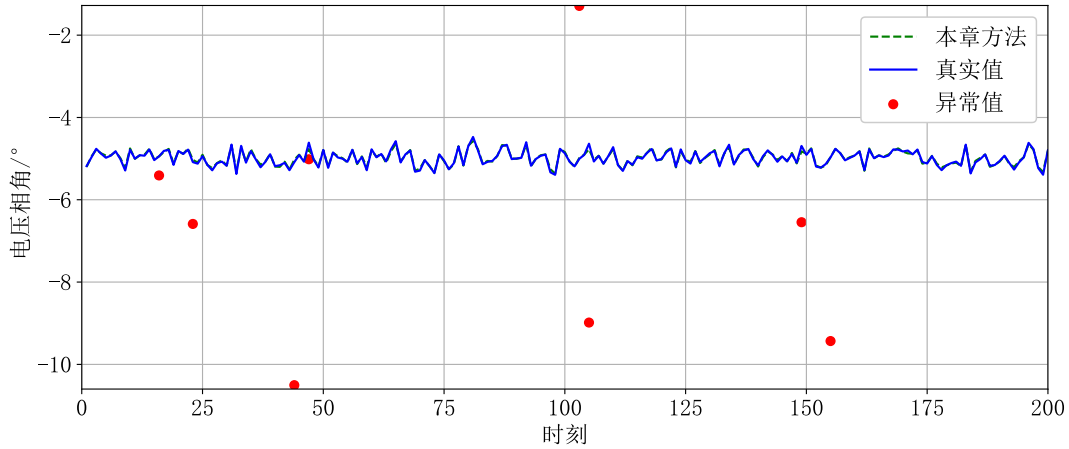


图 4-18 单节点攻击后 AE-WGANGP 电压相角恢复图

图 4-16、4-17、4-18 中节点二分别在 16、23、44、47、103、105、149、155 时刻遭受攻击，其异常值分别为-5.4097、-6.5871、-10.505、-5.0168、-1.2902、-8.9842、-6.5461、-9.433。显然，构造的攻击损害了一定数量的电压相角，使其与攻击前的值有很大的偏差。一旦 FDIA 检测报警，恢复方案将被激活，以恢复攻击前的测量值和状态。比较三个图可以得出，AE-GAN 拟合效果较差，AE 效果略好，而本章采用的方法 AE-WGANGP 效果最好，而被污染的状态值恢复后都与攻击前的值重叠，精度很高。

单节点攻击下，所有节点电压幅值、电压相角的恢复效果分别如表 4-1、4-2 所示：

表 4-1 单节点攻击电压幅值恢复结果

模型	MAE	RMSE	MAPE
AE	0.0013	0.0019	0.1218
AE-GAN	0.0020	0.0028	0.2009
本章方法	**0.0009**	**0.0016**	**0.0904**

由表 4-1 可知，本章方法在上述 MAE、RMSE、MAPE 均为效果最好的。具体而言，其平均绝对误差分别比 AE、AE-GAN 小 0.0004、0.0011，均方根误差分别小 0.0003、0.0012，而平均绝对百分比误差则分别低 3.14%、11.05%。

表 4-2 单节点攻击电压相角恢复结果

模型	MAE	RMSE	MAPE
AE	0.0755	0.1171	0.6359
AE-GAN	0.0844	0.1168	0.7074
本章方法	**0.0454**	**0.0702**	**0.3738**

从表 4-2 中可得，本章方法 AE-WGANGP 的平均绝对误差比 AE、AE-GAN 分别低 0.0301、0.039，AE-GAN 的均方根误差比 AE 低 0.0003，但却比 AE-WGANGP 高出 0.0466。以平均绝对百分比误差而言，AE-WGANGP 比 AE 低 26.21%，比 AE-GAN 低 33.36%。总体而言 AE-WGANGP 效果最优。

2）多节点攻击

在时刻 1 采用多节点攻击，将受损的电压幅值设为零后，送入 AE、AE-GAN、AE-WGANGP，相应的电压幅值恢复结果分别如图 4-19、4-20、4-21 所示。

由图 4-19、4-20、4-21 可知，1 时刻异常节点为 6、7、8 号节点。AE 恢复的 6、8 节点效果较好，7 号节点相对较差。AE-GAN 与 AE-WGANGP 相似，在

三个异常点的恢复效果均较好。除异常点外，AE-GAN 在 4、9、13 节点拟合较差，而 AE-WGANGP 和 AE 在未被攻击的节点依旧能接近真实值。综上，AE-WGANGP 能更好的完成电压幅值的恢复任务。

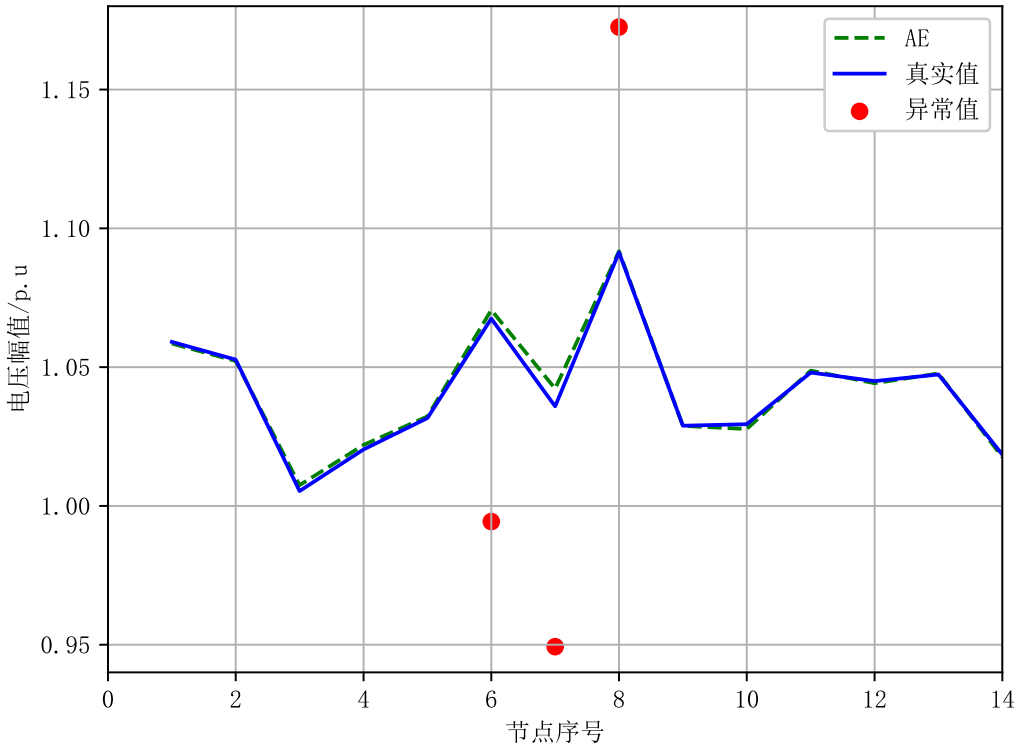


图 4-19 1 时刻多节点攻击后 AE 电压幅值恢复图

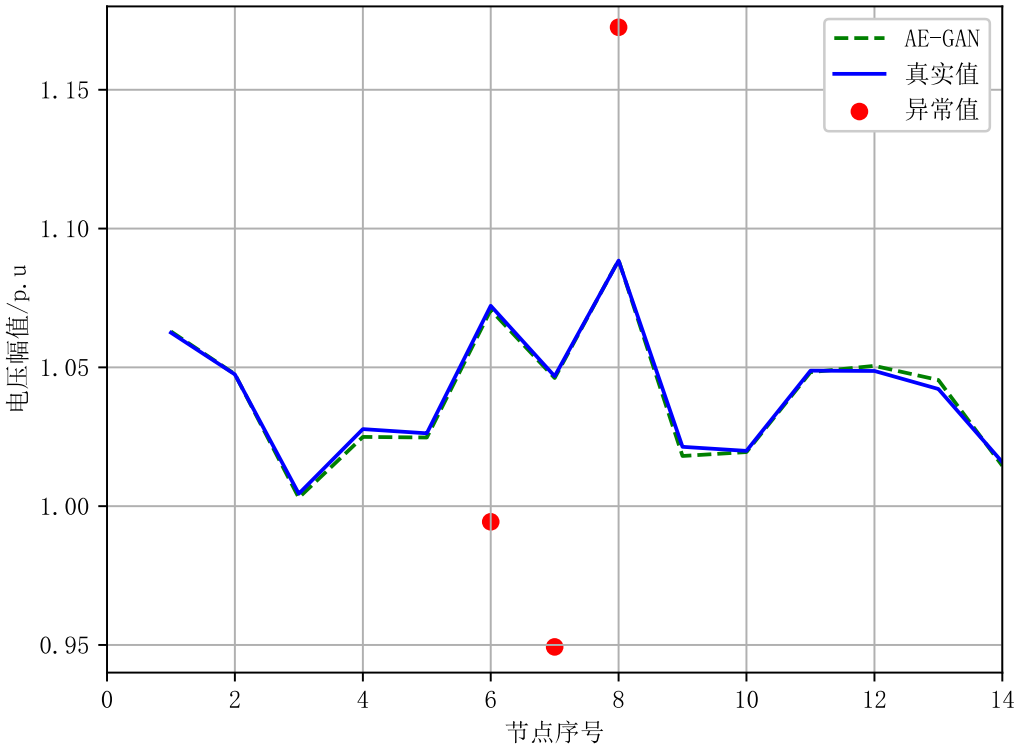


图 4-20 1 时刻多节点攻击后 AE-GAN 电压幅值恢复图

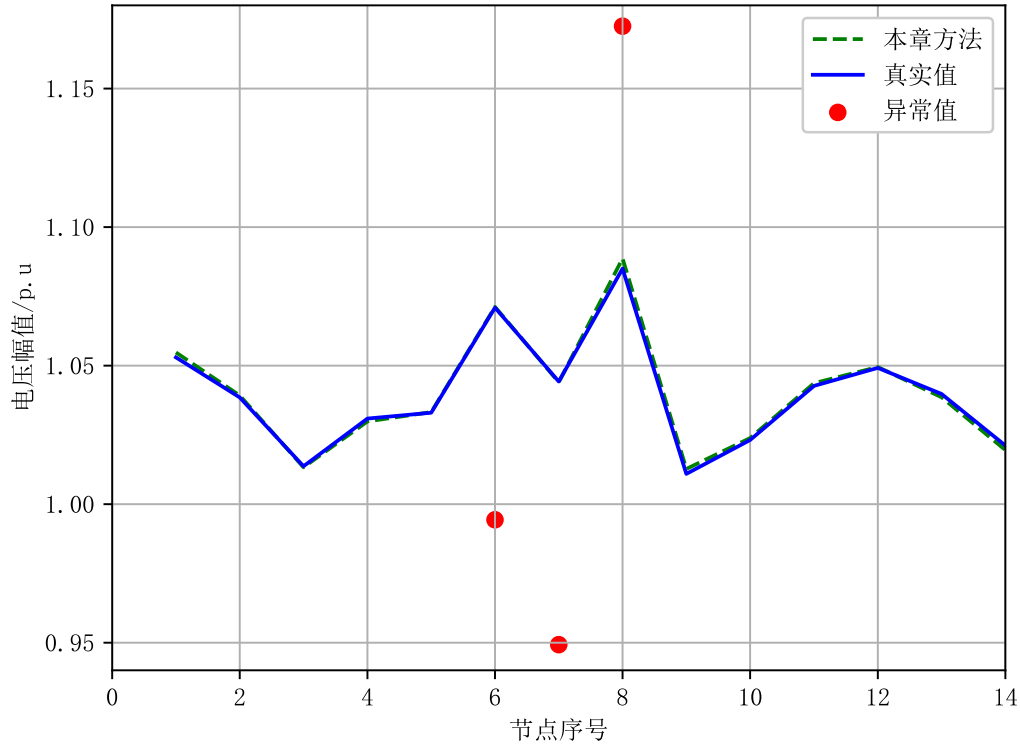


图 4-21 1 时刻多节点攻击后 AE-WGANGP 电压幅值恢复图

在随机多节点攻击后，将被损害节点的数值修改为 0，然后分别送入 AE、AE-GAN、AE-WGANGP 模型，在时刻 1 电压相角恢复的结果如图 4-22 所示：

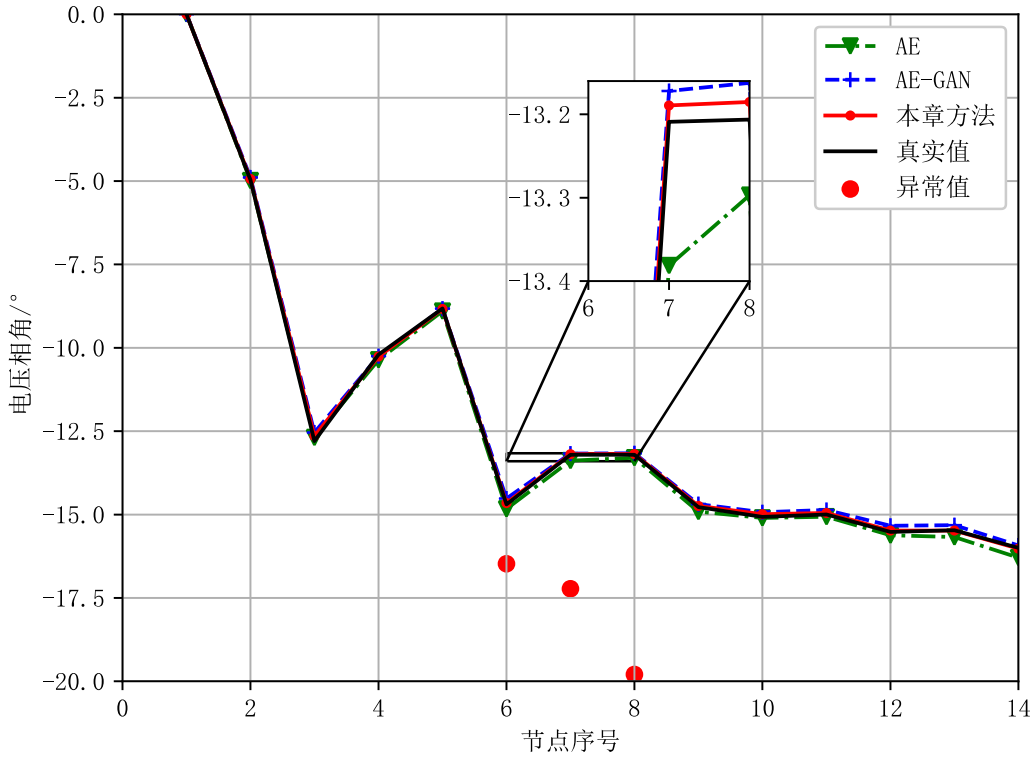


图 4-22 1 时刻多节点攻击后电压相角恢复图

由图 4-22 可知，时刻 1 中异常点为 6、7、8。以节点 7 为例，AE-WGANGP

恢复的后的值更加接近真实值，而 AE-GAN 次之，AE 效果相对较差。

在多节点攻击下，同样以 2 号节点为例展示其电压幅值在 200 个时刻中所受攻击以及恢复的情况，AE、AE-GAN、AE-WGANGP 结果分别如图 4-23、4-24、4-25 所示：

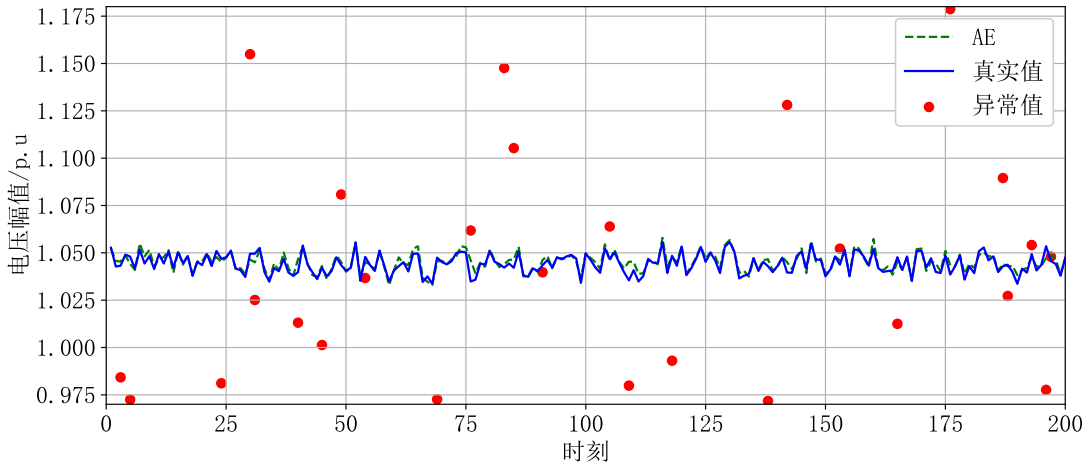


图 4-23 多节点攻击后 AE 电压幅值恢复图

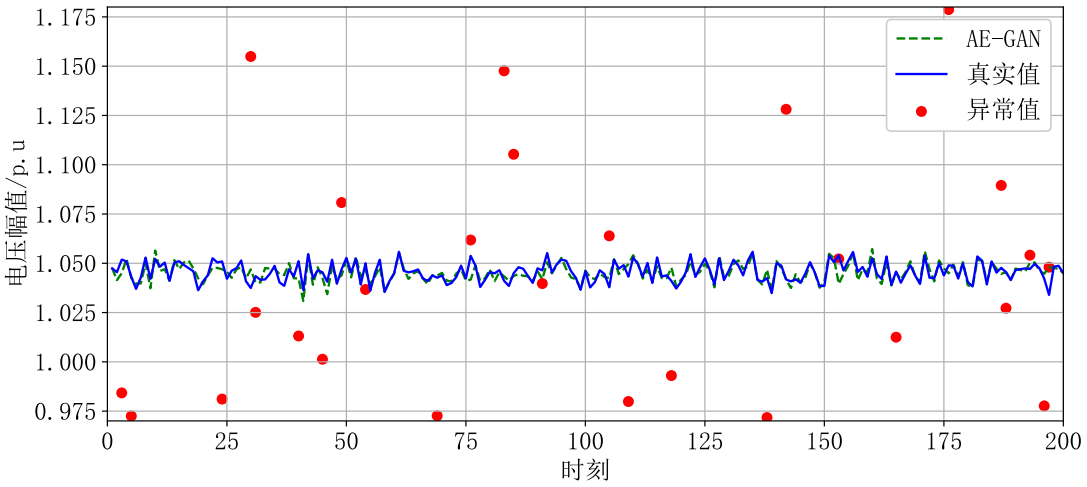


图 4-24 多节点攻击后 AE-GAN 电压幅值恢复图

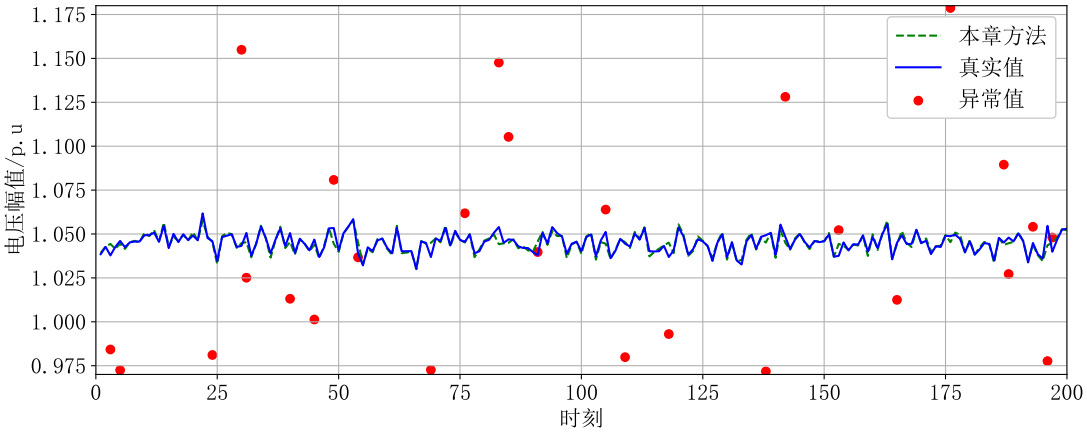


图 4-25 多节点攻击后 AE-WGANGP 电压幅值恢复图

图 4-23、4-24、4-25 中节点二分别在 3、5、24、30、31、40、45、49、54、69、76、83、85、91、105、109、118、138、142、153、165、176、187、188、193、196、197 时刻遭受攻击，其异常值分别为 0.98424、0.97248、0.98109、1.1549、1.0251、1.0131、1.0013、1.0808、1.0367、0.97263、1.0618、1.1476、1.1053、1.0397、1.0639、0.97986、0.99301、0.97174、1.1281、1.0522、1.0125、1.1787、1.0895、1.0272、1.0541、0.97762、1.048。相较于随机单节点攻击，随机多节点攻击明显更能对系统造成威胁，从而影响系统的安全、稳定运行。比较图 4-23、4-24、4-25 可知，不论是 AE、AE-GAN 还是 AE-WGANGP 恢复后的电压幅值基本都能拟合真实值，但也有在部分未被攻击的时刻拟合后的电压幅值略微偏离真实值。

2 号节点其电压相角在 200 个时刻中所受攻击以及 AE、AE-GAN、AE-WGANGP 模型恢复电压相角的效果分别如图 4-26、4-27、4-28 所示：

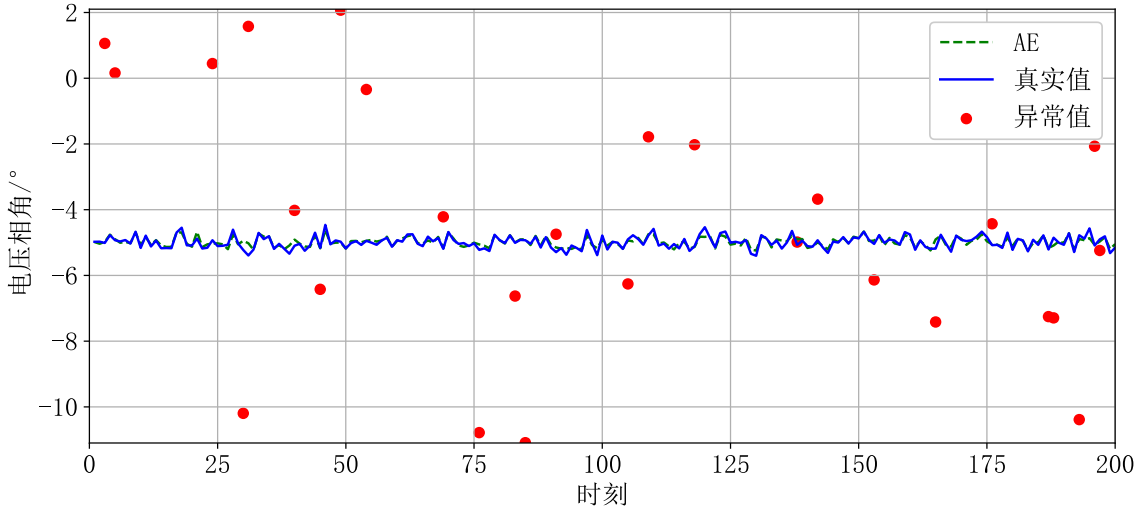


图 4-26 多节点攻击后 AE 电压相角恢复图

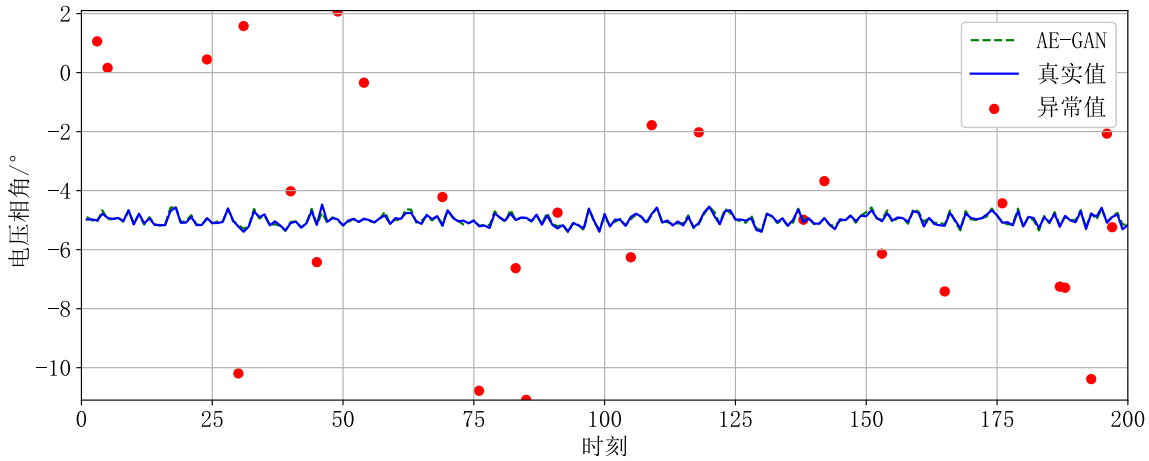


图 4-27 多节点攻击后 AE-GAN 电压相角恢复图

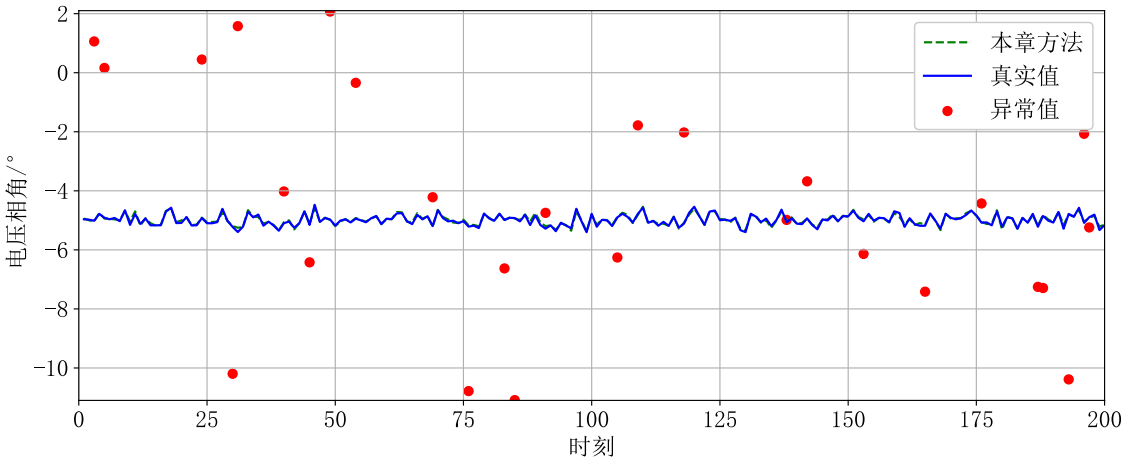


图 4-28 多节点攻击后 AE-WGANGP 电压相角恢复图

在图 4-26、4-27、4-28 中二号节点分别在 3、5、24、30、31、40、45、49、54、69、76、83、85、91、105、109、118、138、142、153、165、176、187、188、193、196、197 时刻被攻击，异常值分别为 1.059、0.15982、0.4458、-10.196、1.5766、-4.0229、-6.4238、2.0694、-0.34405、-4.2172、-10.783、-6.6289、-11.089、-4.747、-6.2586、-1.7847、-2.0246、-4.9863、-3.6793、-6.1378、-7.4168、-4.4291、-7.2542、-7.2927、-10.387、-2.06。

多节点攻击下，所有节点电压幅值的恢复效果如表 4-3 所示：

表 4-3 多节点攻击电压幅值恢复结果

模型	MAE	RMSE	MAPE
AE	0.0022	0.0030	0.2074
AE-GAN	0.0023	0.0033	0.2256
本章方法	**0.0016**	**0.0026**	**0.1534**

由表 4-3 可知，随着被攻击节点的增多，模型恢复效果相对单节点攻击有所下降。总体而言，本章方法 AE-WGANGP 恢复效果比 AE、AE-GAN 略好。具体而言，AE-WGANGP 的平均绝对误差比 AE 低 0.0006、比 AE-GAN 低 0.0007，均方根误差比 AE 低 0.0004、比 AE-GAN 低 0.0007，而平均绝对百分比误差比 AE 低 5.4%、比 AE-GAN 低 7.22%。

多节点攻击下，所有节点电压相角的恢复效果如表 4-4 所示。

从表 4-4 中可得，AE-WGANGP 的平均绝对误差比 AE、AE-GAN 分别低 0.0623、0.0357，其均方根误差相比 AE、AE-GAN 则分别低 0.0929、0.0467，而平均绝对百分比误差相比 AE 低 51.92%，相比 AE-GAN 低 28.58%。综上所述，AE-WGANGP 在数据恢复方面更胜一筹。

表 4-4 多节点攻击电压相角恢复结果

模型	MAE	RMSE	MAPE
AE	0.1302	0.1939	1.0776
AE-GAN	0.1036	0.1477	0.8442
本章方法	**0.0679**	**0.1010**	**0.5584**

4.3 本章小结

针对遭受 FDIA 后电力系统状态数据恢复问题，探究了一种基于 AE-WGANGP 的无监督状态恢复算法来恢复被 FDIA 污染的状态数据。将 AE 与 GAN 结合，采用对抗训练学习数据间的空间关系特征，其无需有标签的历史数据集，只需要提供正常历史数据。而引入的 Wasserstein 距离和梯度惩罚项则可以避免 GAN 在对抗训练中容易出现梯度消失和模式崩溃问题。最终在 IEEE-14 系统上采用随机单节点攻击和随机多节点攻击，并分别使用 AE、AE-GAN、AE-WGANGP 完成数据恢复。数值结果表明，所提出的恢复方案能够准确地将被污染的状态量恢复到攻击前的值，对于 FDIA 污染具有较高的恢复精度。

第 5 章 总结与展望

5.1 总结

FDIA 篡改数据采集和监控系统中的测量值，引起调度中心对当前电网状态的误判，从而达到破坏电力系统安全稳定或获取不法收益的目的，因此对于 FDIA 的检测以及容侵控制研究具有重大意义。本文首先对 Conformer 模型进行改进，利用改进的 Conformer 模型进行虚假数据定位检测；其次，混合三种注意力机制并与堆叠门控循环单元融合，探究 SBIGRU-HA 模型，并利用该模型进行定位检测；最后，结合 AE 与带梯度惩罚的 Wasserstein 生成对抗网络，采用 AE-WGANGP 模型在所生成的两种数据集上分别对电压和电流进行数据恢复，具体内容如下：

1）开展了基于改进 Conformer 的虚假数据注入攻击检测研究。对于 Conformer 模型进行改进，之后基于改进 Conformer 模型在 IEEE-14 以及 IEEE-57 节点系统分别进行定位检测实验，实验证明该算法可以有效检测出 FDIA 的存在。

2）探讨了 SBIGRU 与混合注意力机制的检测模型 SBIGRU-HA。对于 CA、CBAM 进行改进，之后融合 CA、CBAM 和 SimAM 三种注意力机制，再与 SBIGRU 结合，在 IEEE-14 以及 IEEE-57 节点系统分别进行检测实验，实验结果表明，该模型在各项检测指标如准确率、精确率、召回率和 F 分数均效果优秀。

3）研究了虚假数据注入攻击后系统的容侵控制。生成两种数据集：一种包含随机单节点攻击，另一种包含随机多节点攻击。结合 AE 和 WGANGP，采用 AE-WGANGP 在两种攻击下，分别对电压幅值和相角进行数据恢复，实验结果表明，该模型在均方根误差、平均绝对误差和平均绝对百分比误差等指标上均有较好的恢复效果。

5.2 展望

本文所设计的针对虚假数据注入攻击的检测以及容侵控制研究，虽取得了一定成效，但仍有不足之处，如

1）在设计检测模型时，应该考虑训练时间以及检测时间，并且还应该权衡检测准确率、检测时间、模型复杂度和训练时间等多种因素。

2）为了进一步验证检测模型的可靠性和泛化性，在实验中应尽可能站在攻击者的角度，在已知全部拓扑信息、已知部分拓扑信息、仅能攻击有限仪表等多

种情况下，构建虚假数据注入攻击，增加数据集中虚假数据注入攻击的类型。

3）实验考虑的都是离散情况，应该进一步考虑如何在连续情况下完成对智能电网的攻击检测以及容侵控制。

参考文献

[1] Clastres C. Smart grids: Another step towards competition, energy security and climate change objectives[J]. Energy policy, 2011, 39(9): 5399-5408.

[2] Khurana H, Hadley M, Lu N, et al. Smart-grid security issues[J]. IEEE Security & Privacy, 2010, 8(1): 81-85.

[3] Yu X, Xue Y. Smart grids: A cyber–physical systems perspective[J]. Proceedings of the IEEE, 2016, 104(5): 1058-1070.

[4] 张东霞, 苗新, 刘丽平等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.

[5] Fang X, Misra S, Xue G, et al. Smart grid—The new and improved power grid: A survey[J]. IEEE communications surveys & tutorials, 2011, 14(4): 944-980.

[6] Langner R. Stuxnet: Dissecting a cyberwarfare weapon[J]. IEEE Security & Privacy, 2011, 9(3): 49-51.

[7] 郭庆来, 辛蜀骏, 王剑辉等. 由乌克兰停电事件看信息能源系统综合安全评估[J]. 电力系统自动化, 2016, 40(05): 145-147.

[8] 董朝阳, 赵俊华, 文福拴等. 从智能电网到能源互联网: 基本概念与研究框架[J]. 电力系统自动化, 2014, 38(15): 1-11.

[9] Liu Y, Ning P, Reiter M K, et al. False Data Injection Attacks against State Estimation in Electric Power Grids[M]. New York: Assoc Computing Machinery, 2009.

[10] 杨玉泽, 刘文霞, 李承泽等. 面向电力 SCADA 系统的 FDIA 检测方法综述[J]. 中国电机工程学报, 2023, 43(22): 8602-8622.

[11] Musleh A S, Chen G, Dong Z Y. A survey on the detection algorithms for false data injection attacks in smart grids[J]. IEEE Transactions on Smart Grid, 2019, 11(3): 2218-2234.

[12] Xie L, Mo Y, Sinopoli B. Integrity data attacks in power market operations[J]. IEEE Transactions on Smart Grid, 2011, 2(4): 659-666.

[13] 王媛媛. 智能电网中虚假数据注入攻击的检测与防御研究 [D]. 北京: 华北电力大学, 2020.

[14] Yu Z H, Chin W L. Blind false data injection attack using PCA approximation method in smart grid[J]. IEEE Transactions on Smart Grid, 2015, 6(3): 1219-1226.

[15] 田继伟, 王布宏, 尚福特. 基于鲁棒主成分分析的智能电网虚假数据注入攻击[J]. 计算机应用, 2017, 37(7): 1943-1947.

[16] Deng R, Zhuang P, Liang H. False data injection attacks against state estimation in power distribution systems[J]. IEEE Transactions on Smart Grid, 2018, 10(3): 2871-2881.

[17] Zhang J, Chu Z, Sankar L, et al. Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems?[J]. IEEE Transactions on Power Systems, 2018, 33(5): 4775-4786.

[18] Goyel H, Swarup K S. Data Integrity Attack Detection Using Ensemble-Based Learning for Cyber–Physical Power Systems[J]. IEEE Transactions on Smart Grid, 2022, 14(2): 1198-1209.

[19] Yang H, He X, Wang Z, et al. Blind false data injection attacks against state estimation based on matrix reconstruction[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 3174-3187.

[20] Costilla-Enriquez N, Weng Y. Attack power system state estimation by implicitly learning the underlying models[J]. IEEE Transactions on Smart Grid, 2022, 14(1): 649-662.

[21] Chin W L, Lee C H, Jiang T. Blind false data attacks against AC state estimation based on geometric approach in smart grid communications[J]. IEEE Transactions on Smart Grid, 2017, 9(6): 6298-6306.

[22] Chen Y, Huang S, Liu F, et al. Evaluation of reinforcement learning-based false data injection attack to automatic voltage control[J]. IEEE Transactions on Smart Grid, 2018, 10(2): 2158-2169.

[23] Wang Z, Chen Y, Liu F, et al. Power system security under false data injection attacks with exploitation and exploration based on reinforcement learning[J]. IEEE Access, 2018, 6(1): 48785-4896.

[24] Kang J W, Joo I Y, Choi D H. False data injection attacks on contingency analysis: Attack strategies and impact assessment[J]. IEEE Access, 2018, 6(1): 8841-8851.

[25] Li F, Tang Y. False data injection attack for cyber-physical systems with resource constraint[J]. IEEE transactions on cybernetics, 2018, 50(2): 729-738.

[26] Li Z, Shahidehpour M, Alabdulwahab A, et al. Analyzing locally coordinated cyber-physical attacks for undetectable line outages[J]. IEEE Transactions on Smart Grid, 2018, 9(1): 35-47.

[27] Liang G, Weller S R, Luo F, et al. Generalized FDIA-based cyber topology attack with application to the Australian electricity market trading mechanism[J]. IEEE Transactions on Smart Grid, 2017, 9(4): 3820-3829.

[28] Liang G, Weller S R, Zhao J, et al. A framework for cyber-topology attacks: Line-switching and new attack scenarios[J]. IEEE Transactions on Smart Grid, 2019, 10(2): 1704-1712.

[29] Moslemi R, Mesbahi A, Mohammadpour Velni J. Design of robust profitable false data injection attacks in multi-settlement electricity markets[J]. IET generation, transmission & distribution, 2018, 12(6): 1263-1270.

[30] Zhang Z, Deng R, Yau D K Y, et al. Zero-parameter-information data integrity attacks and countermeasures in IoT-based smart grid[J]. IEEE Internet of Things Journal, 2021, 8(8): 6608-6623.

[31] Che L, Liu X, Li Z, et al. False data injection attacks induced sequential outages in power systems[J]. IEEE Transactions on Power Systems, 2018, 34(2): 1513-1523.

[32] Yu L, Sun X M, Sui T. False-data injection attack in electricity generation system subject to actuator saturation: Analysis and design[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(8): 1712-1719.

[33] Liang J, Sankar L, Kosut O. Vulnerability analysis and consequences of false data injection attack on power system state estimation[J]. IEEE Transactions on Power Systems, 2015, 31(5): 3864-3872.

[34] Giani A, Bitar E, Garcia M, et al. Smart grid data integrity attacks[J]. IEEE Transactions on Smart Grid, 2013, 4(3): 1244-1253.

[35] Ahmadian S, Malki H, Han Z. Cyber attacks on smart energy grids using generative adverserial networks[C]//2018 IEEE global conference on signal and information processing (GlobalSIP). IEEE, Anaheim, USA, 26-29 November, 2018: 942-946.

[36] 杨怡, 王勇. 基于 AUKF 的分布式电源系统虚假数据攻击检测方法[J]. 电工电能新技术, 2021, 40(12): 48-55.

[37] 刘鑫蕊, 常鹏, 孙秋野. 基于 XGBoost 和无迹卡尔曼滤波自适应混合预测的电网虚假数据注入攻击检测[J]. 中国电机工程学报, 2021, 41(16): 5462-5476.

[38] Ma C, Liang H, Jing Y. A Novel ZSV-Based Detection Scheme for FDIAs in Multiphase Power Distribution Systems[J]. IEEE Transactions on Smart Grid, 2022, 14(2): 1236-1248.

[39] Wang X, Luo X, Zhang M, et al. Distributed detection and isolation of false data injection attacks in smart grids via nonlinear unknown input observers[J]. International Journal of Electrical Power & Energy Systems, 2019, 110(1): 208-222.

[40] Luo X, Li Y, Wang X, et al. Interval observer-based detection and localization against false data injection attack in smart grids[J]. IEEE Internet of Things Journal, 2020, 8(2): 657-671.

[41] Chakrabarty S, Sikdar B. Detection of malicious command injection attacks on phase shifter control in power systems[J]. IEEE Transactions on Power Systems, 2020, 36(1): 271-280.

[42] Ameli A, Hooshyar A, El-Saadany E F. Development of a cyber-resilient line current differential relay[J]. IEEE Transactions on Industrial Informatics, 2018, 15(1): 305-318.

[43] Wang S, Bi S, Zhang Y J A. Locational detection of the false data injection attack in a smart grid: A multilabel classification approach[J]. IEEE Internet of Things Journal, 2020, 7(9): 8218-8227.

[44] Kurt M N, Ogundijo O, Li C, et al. Online cyber-attack detection in smart grid: A reinforcement learning approach[J]. IEEE Transactions on Smart Grid, 2018, 10(5): 5174-5185.

[45] Goyel H, Swarup K S. Data Integrity Attack Detection Using Ensemble-Based Learning for Cyber–Physical Power Systems[J]. IEEE Transactions on Smart Grid, 2022, 14(2): 1198-1209.

[46] Musleh A S, Chen G, Dong Z Y, et al. Attack detection in automatic generation control systems using lstm-based stacked autoencoders[J]. IEEE Transactions on Industrial Informatics, 2022, 19(1): 153-165.

[47] 李元诚, 曾婧. 基于改进卷积神经网络的电网假数据注入攻击检测方法[J]. 电力系统自动化, 2019, 43(20): 97-104.

[48] Niu X, Li J, Sun J, et al. Dynamic detection of false data injection attack in smart grid using

deep learning[C]//2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, Washington, USA, 18-21 February, 2019: 1-6.

[49] Zhang Y, Wang J, Chen B. Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach[J]. IEEE Transactions on Smart Grid, 2020, 12(1): 623-634.

[50] Wu T, Xue W, Wang H, et al. Extreme learning machine-based state reconstruction for automatic attack filtering in cyber physical power system[J]. IEEE Transactions on Industrial Informatics, 2020, 17(3): 1892-1904.

[51] Adhikari U, Morris T H, Pan S. Applying non-nested generalized exemplars classification for cyber-power event and intrusion detection[J]. IEEE Transactions on Smart Grid, 2018, 9(5): 3928-3941.

[52] Adhikari U, Morris T H, Pan S. Applying hoeffding adaptive trees for real-time cyber-power event and intrusion classification[J]. IEEE Transactions on Smart Grid, 2018, 9(5): 4049-4060.

[53] Pan S, Morris T, Adhikari U. Classification of disturbances and cyber-attacks in power systems using heterogeneous time-synchronized data[J]. IEEE Transactions on Industrial Informatics, 2015, 11(3): 650-662.

[54] Li Y, Wei X, Li Y, et al. Detection of false data injection attacks in smart grid: A secure federated deep learning approach[J]. IEEE Transactions on Smart Grid, 2022, 13(6): 4862-4872.

[55] Huang K, Xiang Z, Deng W, et al. False data injection attacks detection in smart grid: A structural sparse matrix separation method[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(3): 2545-2558.

[56] Chakrabarty S, Sikdar B. Unified Detection of Attacks Involving Injection of False Control Commands and Measurements in Transmission Systems of Smart Grids[J]. IEEE Transactions on Smart Grid, 2021, 13(2): 1598-1610.

[57] Liu X K, Wen C, Xu Q, et al. Resilient control and analysis for DC microgrid system under DoS and impulsive FDI attacks[J]. IEEE Transactions on Smart Grid, 2021, 12(5): 3742-3754.

[58] Jiang Y, Yang Y, Tan S C, et al. Distributed sliding mode observer-based secondary control for DC microgrids under cyber-attacks[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2020, 11(1): 144-154.

[59] Zuo S, Altun T, Lewis F L, et al. Distributed resilient secondary control of DC microgrids against unbounded attacks[J]. IEEE Transactions on Smart Grid, 2020, 11(5): 3850-3859.

[60] Li J, Yang D F, Gao Y C, et al. An adaptive sliding-mode resilient control strategy in smart grid under mixed attacks[J]. IET Control Theory & Applications, 2021, 15(15): 1971-1986.

[61] Sargolzaei A, Yazdani K, Abbaspour A, et al. Detection and mitigation of false data injection attacks in networked control systems[J]. IEEE Transactions on Industrial Informatics, 2019,

16(6): 4281-4292.

[62] Abhinav S, Modares H, Lewis F L, et al. Resilient cooperative control of DC microgrids[J]. IEEE Transactions on Smart Grid, 2018, 10(1): 1083-1085.

[63] Habibi M R, Baghaee H R, Dragičević T, et al. False data injection cyber-attacks mitigation in parallel DC/DC converters based on artificial neural networks[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020, 68(2): 717-721.

[64] Rajabi A, Bobba R B. Resilience against data manipulation in distributed synchrophasor-based mode estimation[J]. IEEE Transactions on Smart Grid, 2021, 12(4): 3538-3547.

[65] Cui J, Gao B, Guo B. A novel detection and defense mechanism against false data injection attack in smart grids[J]. IET Generation, Transmission & Distribution, 2023, 17(20): 4514-4524.

[66] Huang X, Qin Z, Xie M, et al. Defense of massive false data injection attack via sparse attack points considering uncertain topological changes[J]. Journal of Modern Power Systems and Clean Energy, 2021, 10(6): 1588-1598.

[67] Li Y, Wang Y, Hu S. Online generative adversary network based measurement recovery in false data injection attacks: A cyber-physical approach[J]. IEEE Transactions on Industrial Informatics, 2019, 16(3): 2031-2043.

[68] Wang H, Wen X, Xu Y, et al. Operating state reconstruction in cyber physical smart grid for automatic attack filtering[J]. IEEE Transactions on Industrial Informatics, 2020, 18(5): 2909-2922.

[69] Jorjani M, Seifi H, Varjani A Y, et al. An optimization-based approach to recover the detected attacked grid variables after false data injection attack[J]. IEEE Transactions on Smart Grid, 2021, 12(6): 5322-5334.

[70] 吴新辉, 毛政元, 翁谦等. 利用基于残差多注意力和 ACON 激活函数的神经网络提取建筑物[J]. 地球信息科学学报, 2022, 24(04): 792-801.

[71] 马晓东, 魏利胜, 刘小珲. 基于新型 YOLOv5 算法的磁悬浮球精确识别[J]. 电子测量与仪器学报, 2022, 36(08): 204-212.

[72] Shang R, Ren J, Zhu S, et al. Hyperspectral image classification based on pyramid coordinate attention and weighted self-distillation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60(1): 1-16.

[73] 王宁波, 王先培. 容侵技术在电力系统数据网络安全中的应用[J]. 电力自动化设备, 2004, (10): 35-38.

[74] Kim C, Park S, Hwang H J. Local stability of wasserstein gans with abstract gradient penalty[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(9): 4527-4537.

攻读学位期间发表的学术论文

[1] ***Wang Ruiren***, Wei Lisheng, Zhang Pinggai. Research on False Data Injection Attack in Smart Grid Based on Improved Conformer[C]//2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, HeFei, China, 27-29 August, 2023: 1318-22. (EI: 20240815614510)

[2] Liu Rui, Wei Lisheng, ***Wang Ruiren***. Brief-term Electricity Load Forecasting Based on PSO-optimized BP Neural Network[C]//2023 3nd International Conference on Computer, Control and Robotics (ICCCR). IEEE, ShangHai, China, 24-26 March, 2023: 404-08. (EI: 20233414619610)

[3] Wei Lisheng, ***Wang Ruiren***, Zhang Qian, et al. Two-Layer Defense Architecture for False Data Injection Attack in Grid CPS[C]//2024 36th Chinese Control and Decision Conference (CCDC). IEEE, Xi'An, China, 25-27 May, 2024. (已录用，EI，论文编号：0667)

[4] ***王瑞仁***, 魏利胜. 融合 SBIGRU 与注意力机制的虚假数据注入攻击检测[J]. 重庆工商大学学报(自然科学版), 2024. (已录用，三类)

攻读学位期间获得的奖励

[1] 获得 2023-2024 学年校研究生**二等奖学金**

[2] 获得 2022-2023 学年校研究生**二等奖学金**

[3] 获得 2021-2022 学年校研究生**二等奖学金**

[4] 获得 2022 年校第八届"互联网+"大学生创新创业大赛研究生创意组**三等奖**

致　谢

首先，我要由衷感谢我的导师尊敬的魏利胜教授以及校外导师郭春淼老师。导师不仅在学术上给予我耐心的指导和教诲，还在生活上给予我无私的关怀和鼓励。其严谨求实的治学态度和对学术的热情激励着我不断进步和探索，在整个研究过程中，导师始终给予我宝贵的建议和指导，使我受益匪浅。导师的学识渊博、风趣幽默的教学方式深深熏陶着我，让我倍感荣幸能有这样优秀的导师指引我的学术之路。导师还十分关心学生的就业情况，感谢导师在我找工作中给我提供的帮助和指导。

其次，感谢我的师兄王宁、谷峥岩、杨奔奔、沈鑫、汪瑞、马晓东、师姐张倩，在学习上对我的指导；感谢我的同门刘瑞、唐绍语、高港、项诗雨，在逆境中给我帮助和鼓励；感谢我的师弟陶新杰、朱圣博、邵自强、武涛、师妹李明、王勤、金莉莉以及大实验室的其他小伙伴的陪伴；感谢我的室友肖建、姚满宇、张兴旺，感谢你们为我提供良好的睡眠环境；感谢对面宿舍的胡宇杨陪我切磋台球，你永不服输的精神永远激励着我。

同时，感谢我的父母这么多年以来为我保驾护航，我才能无后顾之忧的安心学习，当我毕业你们头发也已花白，未来看我的；感谢我的姐姐、姐夫在生活上给予我的帮助以及可爱的侄子给我带来的欢乐。再次由衷的感谢每一位帮助我，关心我，支持我，鼓励我，陪伴我的人，谢谢大家！

最后，对评阅本论文的各位教授和专家致以真诚的感谢！

王瑞仁

安徽工程大学电气工程学院

二０二四年五月

尚德教学 唯实惟新