

# 基于深度学习的智能物联网时序数据异常检测

## 本文贡献

- 阐述基于GAN的异常检测及轻量化算法流程
- 针对智能物联网产生的**高维时序数据**，提出了基于GAN且包含注意力机制的异常检测模型**Att-ADGAN (Attention-Anomaly Detection GAN)**
- 针对计算及存储**资源受限**的边缘设备，在Att-ADGAN的基础上，提出了**两阶段知识蒸馏框架SKDGAN**

## 基于GAN的异常检测模型Att-ADGAN

### 模型架构

异常检测模型Att-ADGAN如图所示：

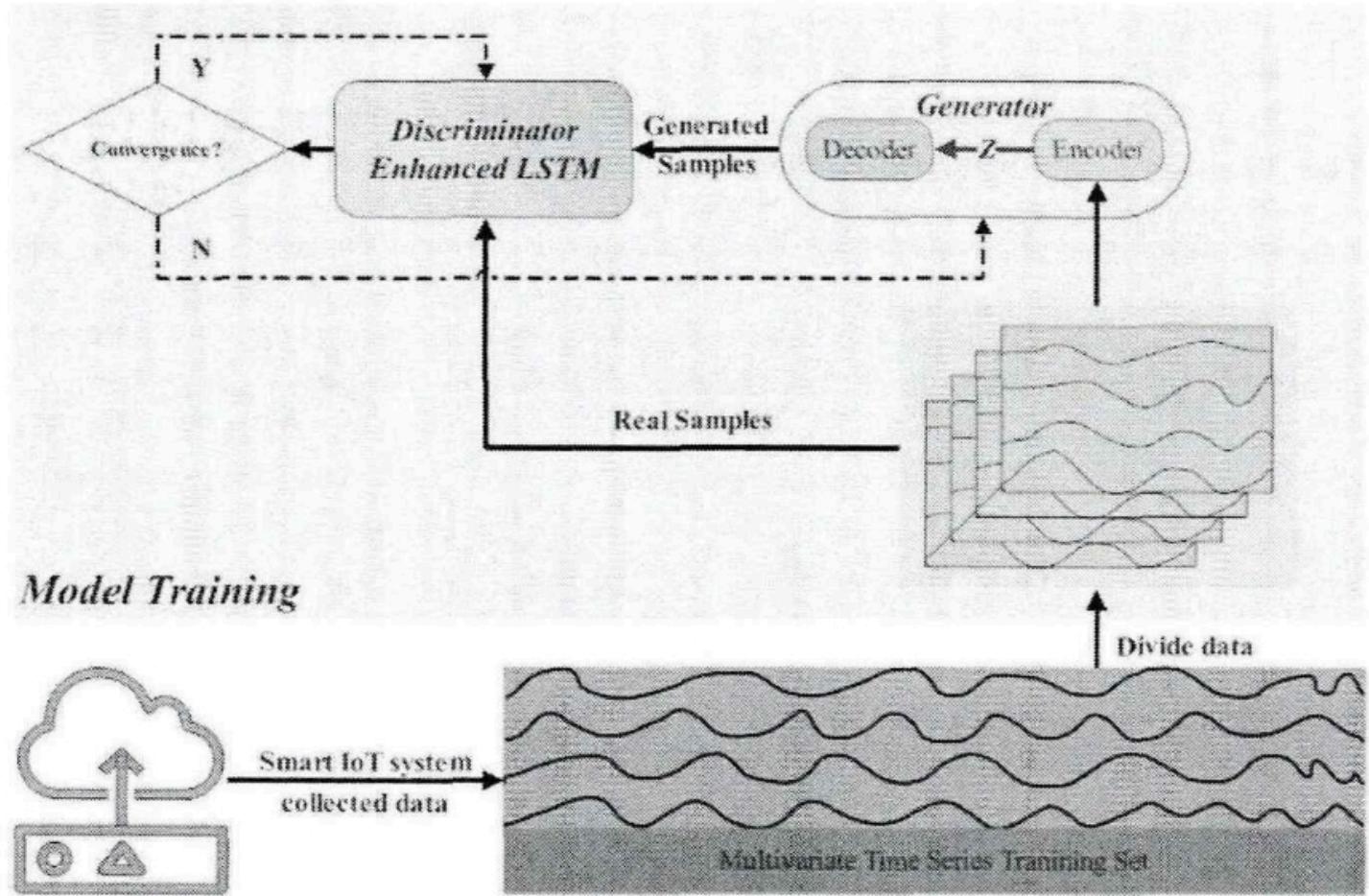


图 3-1 Att-ADGAN 模型总体结构

Att-ADGAN 使用 **LSTM 作为生成器和判别器** 的基本模型来处理复杂的多维时间序列数据。

对于多维时间序列数据，将数据划分为**子序列**，子序列通过滑动窗口机制被送到模型中，窗口大小设置为：

$$s_w = 30 \times i, i = 1, 2, \dots, n$$

模型训练步骤

- 1. 数据预处理
  - 1) 将数据集划分为训练集、验证集和测试集
  - 2) 其中**训练集需要全部为正常数据**，确保模型准确学习到正常数据的分布模式
  - 3) 训练集和验证集被大小为  $s_w$  的滑动窗口划分为统一的子序列，需要对验证数据集进行标记
- 2. 模型训练

**生成器试图通过编码器-解码器架构生成类真实样本欺骗判别器，判别器将尽可能区分出生成样本与真实样本**

1) 通过GAN模型学习数据

分别**对抗学习** 两个映射函数  $\varepsilon : X \rightarrow Z$  和  $G : Z \rightarrow X$  (其中  $X$  是训练样本，  $Z$  是潜在空间向量)

通过两个映射函数可以实现数据重构  $x_i \rightarrow \varepsilon(x_i) \rightarrow G(\varepsilon(x_i)) \approx x_i$

## 2) 训练细节

为确保模型学习到正态数据的分布模式。**训练阶段输入数据全是正常数据**

将生成器的**输出**  $G(\varepsilon(x_i))$  和 **原始数据**  $x_i$  发送到判别器进行训练

## 3) 损失函数定义

通过G和D的对抗训练不断提高其性能，直到达到设定的迭代次数或模型收敛

\*\*对抗损失：\*\*生成器会尽量减少损失，判别器试图最大化损失

$$L_{adv} = E_{X \sim p_X} [\log(D(X))] + E_{X \sim p_X} [\log(1 - D(\varepsilon(X)))] \quad (1)$$

注：其中 $D(X)$ 是判别器输出， $E_{X \sim p_X}$ 表示从实空间采样的真实样本， $\log(D(X))$ 表示判别器预期原始样本为真， $\log(1 - D(\varepsilon(X)))$ 表示预期生成的样本为假。

\*\*特征损失：\*\*使用L2范式

$$L_{fea} = E_{X \sim p_X} \|f(X) - f(G(\varepsilon(X)))\|_2 \quad (2)$$

注：其中 $f(\cdot)$ 是判别器最后一层输出，损失是 $f(X)$ 和 $f(G(\varepsilon(X)))$ 的L2范数。

**映射损失：**为确保原始数据 $x_i$ 可以映射到潜空间 $z_i$ ，最小化原始与重构样本的残差的L2范数

$$L_{map} = E_{X \sim p_X} \|X - G(\varepsilon(X))\|_2 \quad (3)$$

## 总损失：

$$L_G = \lambda_a L_{adv} + \lambda_f L_{fea} + \lambda_m L_{map} \quad (4)$$

注： $\lambda_a, \lambda_f$  和  $\lambda_m$  表示权重

## • 3. 生成器与判别器

为提高重构效果，将生成器和判别器的**基本模型改进为Enhanced LSTM结构**，Enhanced LSTM结构如下所示：

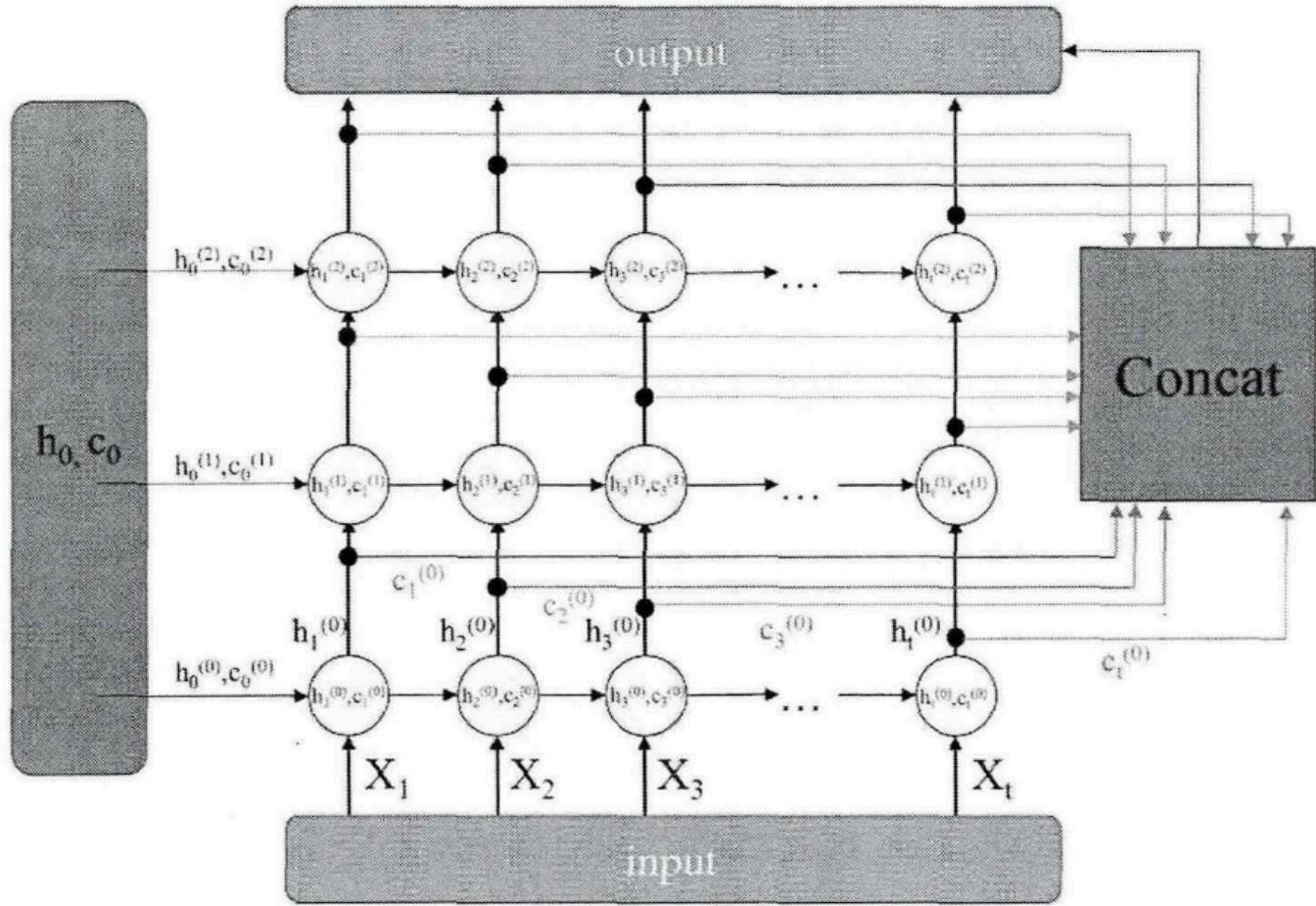


图 3-2 Enhanced LSTM 结构

在编码器和解码器**之前连接一个注意力模块**, 结构如下所示:

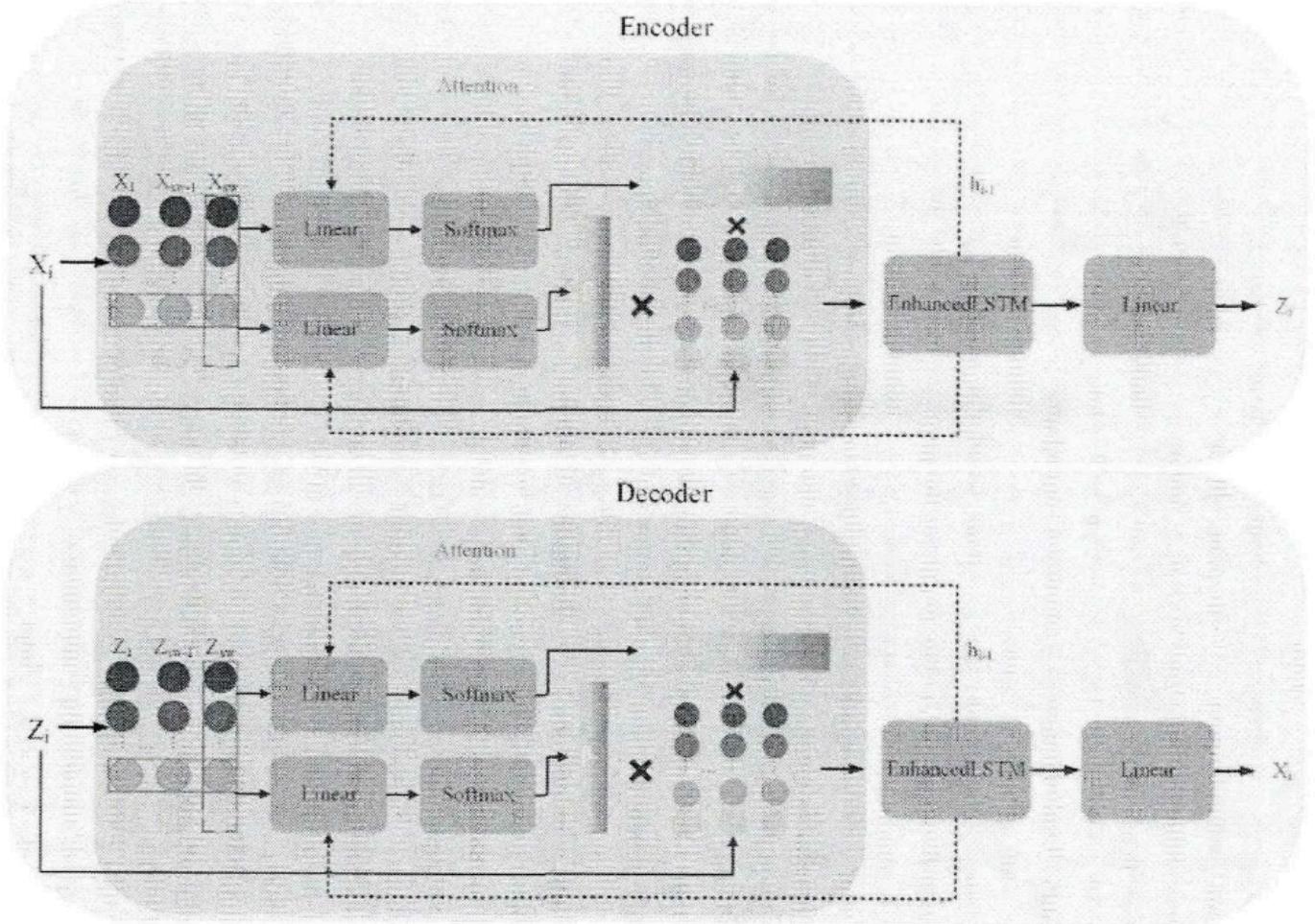


图 3-3 编码器及解码器内部结构

第一阶段，在**时间维度**上执行注意力计算，时间维度的重要性 $a_{ij}$ 通过式(5)和(6)计算，如下：

$$S_i^{time} = \tanh(W_h^a \cdot h_{i-1} + X_i^T w_x^a + b^a) \quad (5)$$

$$a_{ij} = \frac{\exp(s_{ij}^{time})}{\sum_{j'=1}^{S_w} \exp(s_{ij'}^{time})}, j = 1, 2, \dots, S_w \quad (6)$$

第二阶段，在**特征维度**上执行注意力计算，特征维度的重要性 $b_{ij}$ 通过式(5)和(6)计算，如下：

$$S_i^{feature} = \tanh(W_h^\beta \cdot h_{i-1} + X_i^T w_x^\beta + b^\beta) \quad (7)$$

$$b_{ij} = \frac{\exp(s_{ij}^{feature})}{\sum_{k'=1}^d \exp(s_{ik'}^{feature})}, k = 1, 2, \dots, d \quad (8)$$

经过注意力机制，提高输入的**特征和时间相关性**，之后在输入到Enhanced LSTM中。

## 异常检测

异常检测过程如图所示：

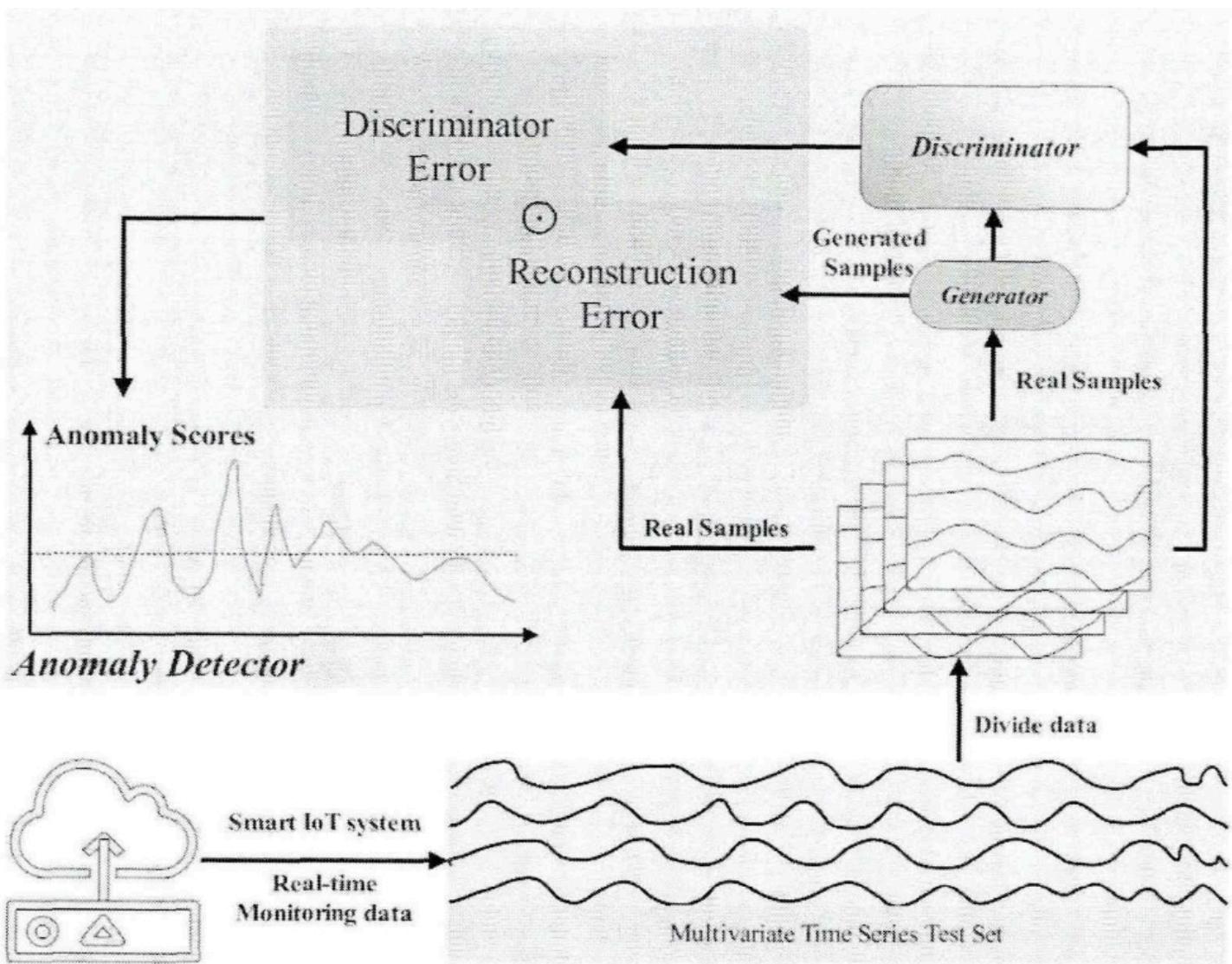


图 3-4 异常检测模型

- 1. 异常检测细节

使用与训练集相同的数据预处理方法，根据时间窗口将标记测试集划分为子序列。

将驻点误差与曲线相似度组合为最终的异常分数。

驻点误差：

$$l_d = \sum_{i=1}^n |x_t^{test,i} - G(\varepsilon(x_t^{test,i}))| \quad (9)$$

注:  $x_t^{test,i} \in R^n$  为 t 时刻第 i 个变量的测量值

曲线相似度 (使用DTW算法) :

$$S_t = W^* = DTW(X, \hat{X}) = \min \left[ \frac{1}{k} \sqrt{\sum_{k=1}^k w_k} \right] \quad (10)$$

最终重构误差:  $L_R = \alpha L_d + \beta S_t$

- 2. 实验部分

使用数据集: **SWMRU、KDDCup99、HomeC**

评估指标: **精度 (Precision) 、召回率 (Recall) 和 F1 分数**

$$Pre = \frac{TP}{TP + FP}$$

$$Rec = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec}$$

注: TP (True Positives) : 是正确检测到异常

FP (False Positives) 是错误检测到正常

TN (True Negatives) 是正确检测到正常

FN (False Negatives) 是错误检测到正常

- 3. 结果与讨论

1) 数据重构性能

实验证明加入 attention 机制重构更加准确, 使用最大平均差异 (Maximum Mean Discrepancy, MMD) 进行评估

2) 窗口设置与重构误差度量值

证明窗口大小对实验结果有影响

# 基于Att-ADGAN的知识蒸馏框架S-KDGAN

为解决资源受限设备提出了一种用于高维时间序列数据的两阶段知识蒸馏框架S-KDGAN

本文使用**面向过程**的知识蒸馏框架S-KDGAN，该框架使用模型架构及参数完整的Att-ADGAN作为教师网络，通过中间层信息何输出信息来指导轻量化的学生网络。蒸馏整体结构图如下：

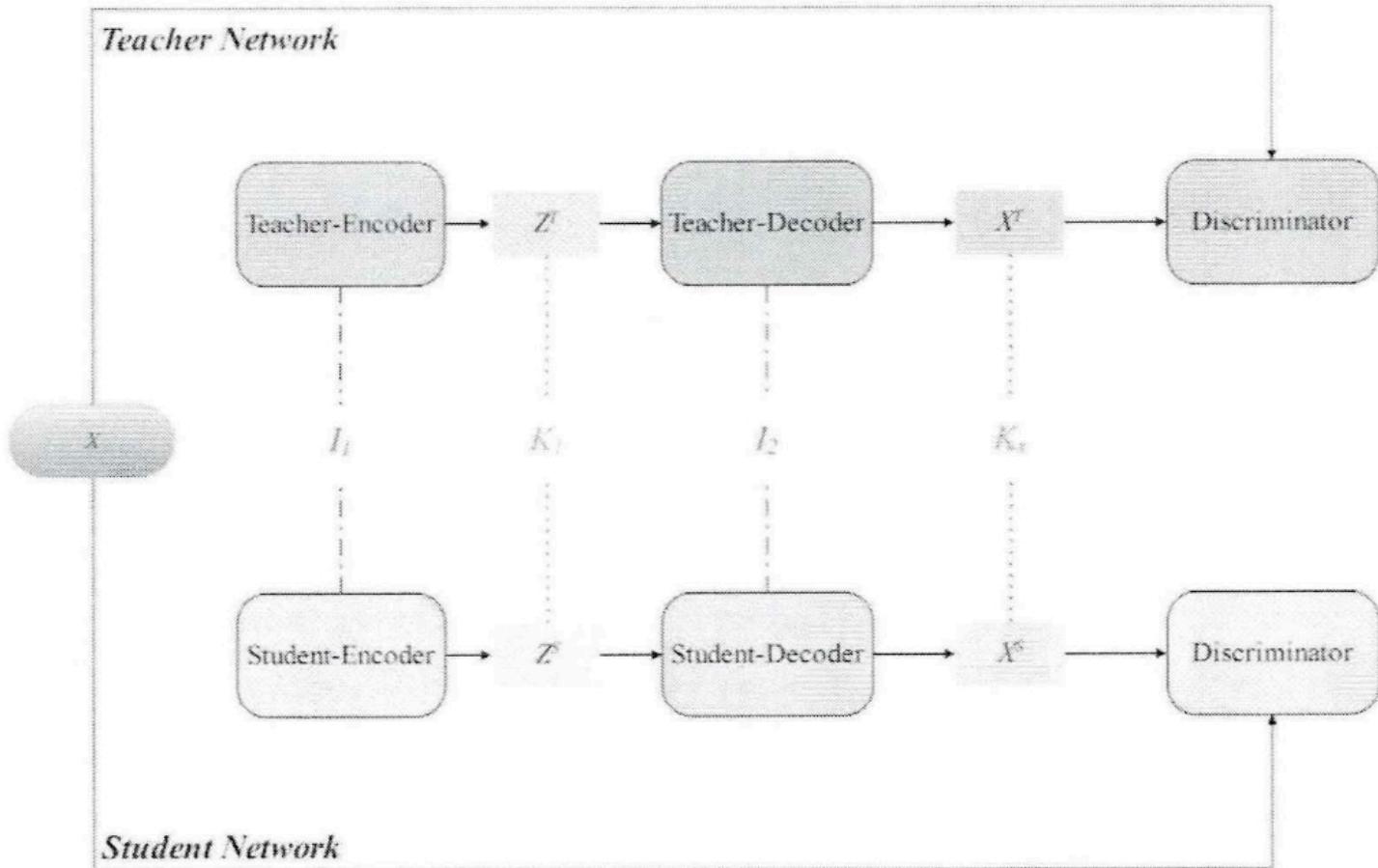


图 4-1 S-KDGAN 蒸馏模型总览

## 蒸馏损失

教师网络的生成器损失函数定义：

**对抗损失：**

$$L_{adv}^T = E_{X \sim p_X} [\log(D(X))] + E_{X \sim p_X} [\log(1 - D(G(\varepsilon(X)))))] \quad (4-1)$$

**特征损失：**

$$L_{fea}^T = E_{X \sim p_X} \|f(X) - f(G(\varepsilon(X)))\|_2 \quad (4-2)$$

**映射损失：**

$$L_{map}^T = E_{X \sim p_X} \|X - G(\varepsilon(X))\|_2 \quad (4-3)$$

**生成器总损失:**

$$L_G^T = \lambda_a^T L_{adv}^T + \lambda_f^T L_{fea}^T + \lambda_m^T L_{map}^T \quad (4-4)$$

教师网络的判别器损失函数定义:

**判别器总损失:**

$$L_D^T = E_{X \sim p_X} [\log(D(X))] + E_{X \sim p_X} [\log(1 - D(G(\varepsilon(X))))] \quad (4-5)$$

注:  $\lambda_a^T$ 、 $\lambda_f^T$ 和 $\lambda_m^T$ 是其损失对应的加权参数,  $L_G^T$ 是生成器对应的损失函数,  $L_D^T$ 是判别器对应的损失函数。

因为学生模型与教师模型目标一致, 因此学生成器  $L_G^T$  和判别器损失  $L_D^T$  与教师网络一致。

为训练蒸馏网络, 设计了四个损失以测量向量间的相似性, 编码器中间层损失  $I_1$ , 潜向量输出损失  $K_1$ , 判别器中间层损失  $I_2$ , 重构损失  $K_2$ 。

细节: 损失由L1距离给出, L2距离会导致模型误差变大。分别计算教师网络与学生网络编码器中间层损失  $I_1$ , 潜向量输出损失  $K_1$ , 判别器中间层损失  $I_2$ , 重构损失  $K_2$  的L1距离。

**蒸馏损失:**

$$K_d = W_1 I_1 + \gamma_1 K_1 + W_2 I_2 + \gamma_2 K_2 \quad (4-6)$$

S-KDGAN训练一共包含5个损失函数: 教师网络生成器损失  $L_G^T$ , 教师网络判别器损失  $L_D^T$ , 学生网络生成器损失  $L_G^S$ , 学生网络判别器损失  $L_D^S$ , 蒸馏损失  $K_d$ , 为研究是否与学生网络结合进行协同训练也将影响最终的蒸馏结构。设计了四个关于上述损失的组合:

**KD-A:**  $\mathcal{L}_1 = \{K_d\}$

注: 教师网络与学生网络的损失函数均不参与训练, 训练过程仅依靠蒸馏损失  $K_d$

**KD-B:**  $\mathcal{L}_2 = \{L_G^S, L_D^S, k_d\}$

注: 教师网络的损失函数不参与训练, 效果较KD-A会有所提示。

**KD-C:**  $\mathcal{L}_3 = \{L_G^T, L_D^T, k_d\}$

注: 学生网络的损失函数不参与训练, 较KD-B参数量有所提升, 预期蒸馏效果高于KD-B.

**KD-D:**  $\mathcal{L}_4 = \{L_G^T, L_D^T, L_G^S, L_D^S, k_d\}$

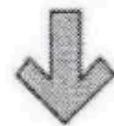
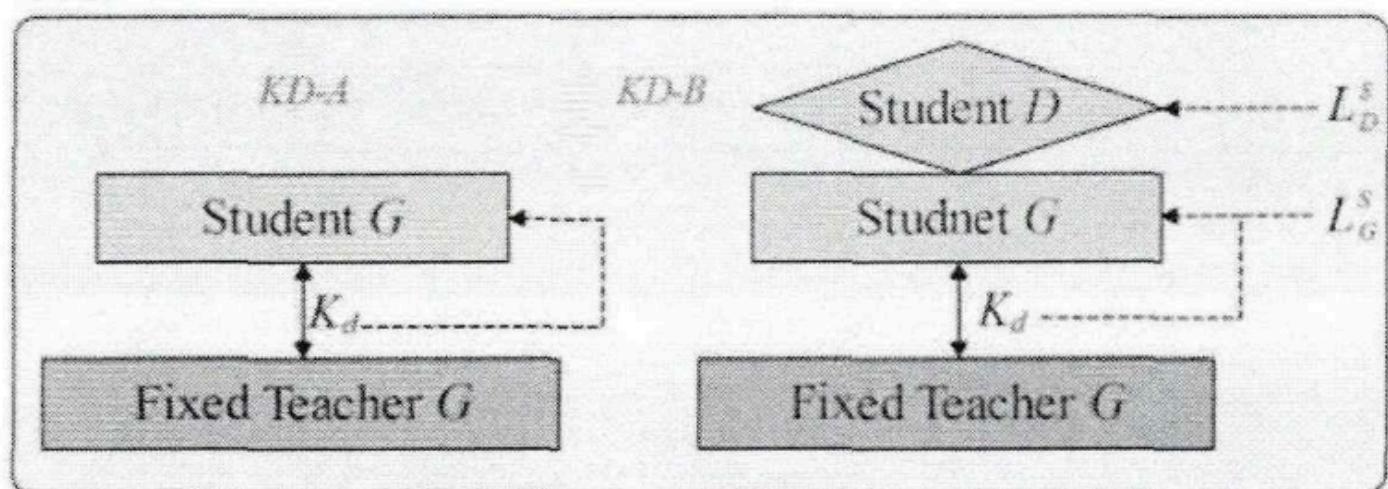
注: 所有损失函数均参与训练

通过四种组合, 试图找出最佳的组合训练方案。根据上述四种蒸馏结构, 提出了一个两阶段的训练方式

## 二阶段蒸馏模型

S-KDGAN通过两阶段的训练方式不断提高学生网络的蒸馏效果，两阶段蒸馏示意图如下：

First



Second

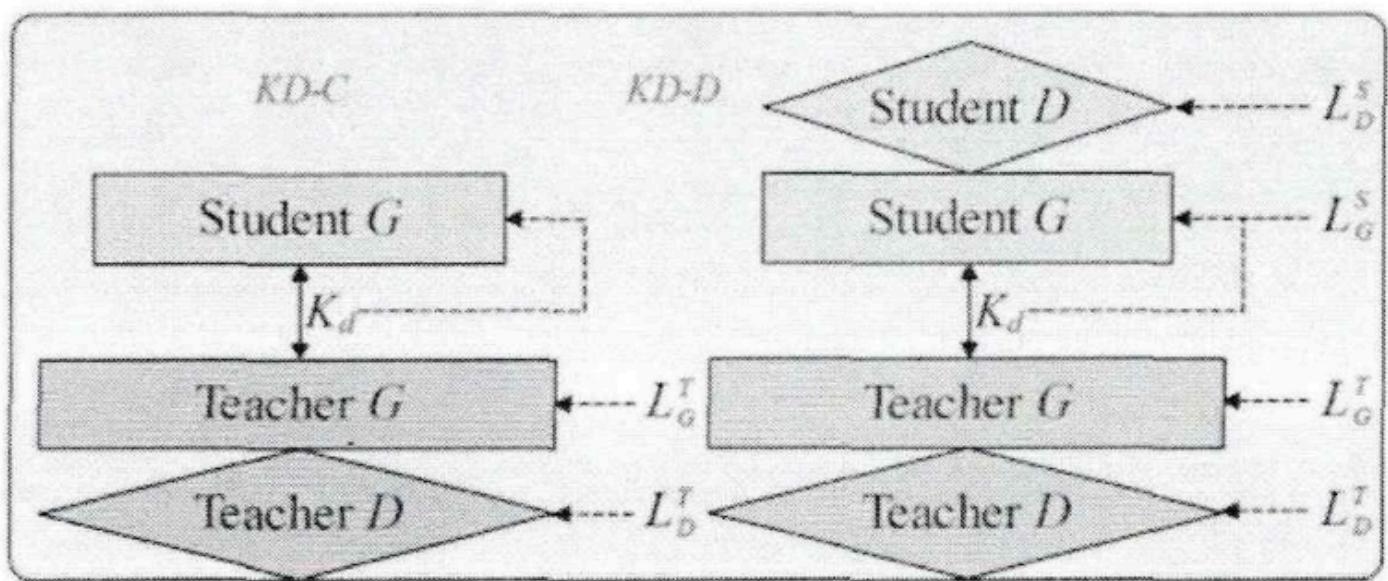


图 4-3 S-KDGAN 的两阶段蒸馏过程

## 模型详细参数

**教师模型：** 教师网络遵循Att-ADGAN的全部网络架构，其中训练阶段，步长设置为10，测试阶段步长为时间窗口大小，Enhanced LSTM深度设置为5，隐藏单元设置为100，潜向量维度设置为15，一个epoch训练判别器一次，生成器三次，epoch设置为500

**学生模型：** 学生网络需要进行轻量化设置，去除了注意力机制，使用普通LSTM单元。LSTM单元深度设置为1，隐藏单元设置为50，潜向量维度设置为5

注：由于学生网络和教师网络的潜在向量维度不一致，所以在学生网络编码器输出后会经过一个线性层。