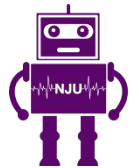




南京大學
NANJING UNIVERSITY



RINC

Robotic Intelligence & Neural Computing Group

具有生物合理性的脉冲神经网络 局部非 BP 训练算法研究

答辩人：胡嘉骏 502022370015

导师：申富饶 教授

日期：2025年5月6日

誠樸雄偉 勵學敦行

壹 研究背景

貳 研究内容

- 基于混合损失局部学习的脉冲神经网络训练算法
- 基于HSIC瓶颈的脉冲神经网络训练算法

叁 实际应用

肆 总结

目录

第一部分

研究背景

Research Background

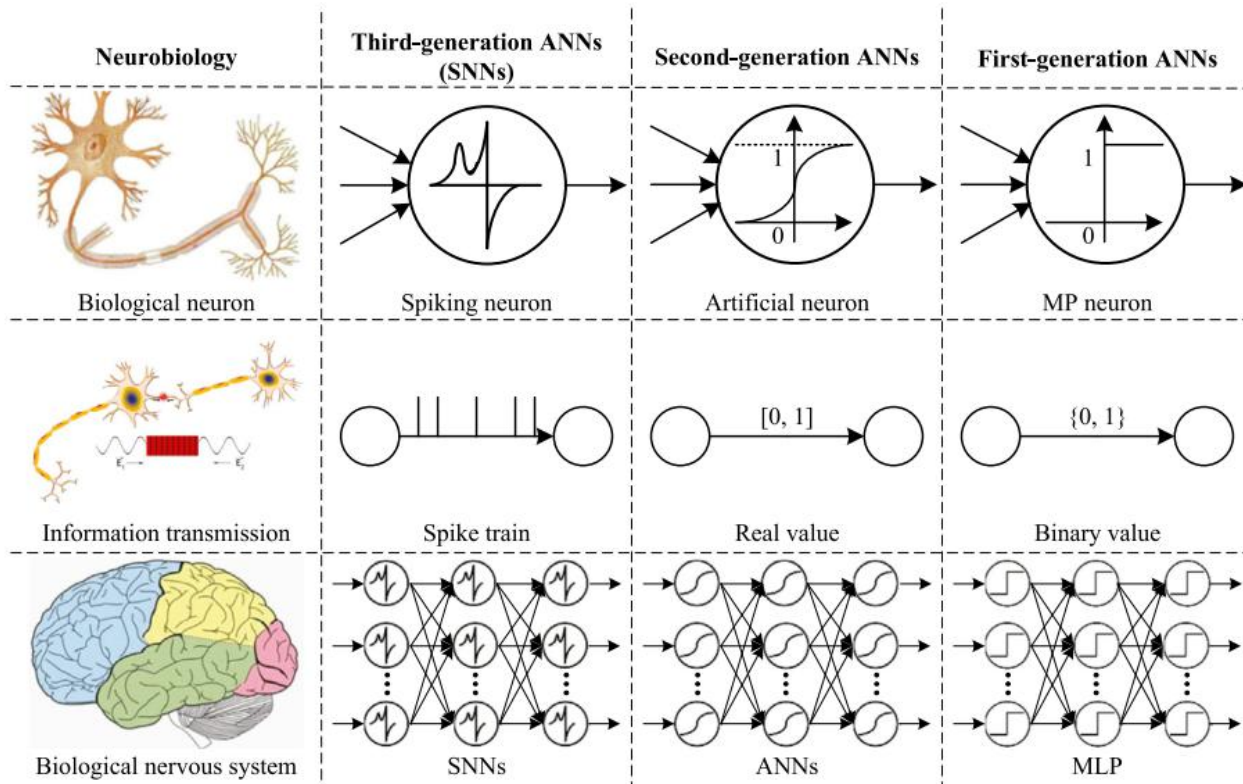
背景简介 | 研究意义 | 困难与挑战

誠樸雄偉 勵學敦行

1.1 背景简介：脉冲神经网络

脉冲神经网络（SNN）被誉为**第三代神经网络**，是连接神经科学和人工智能的桥梁，有以下特点

- 事件驱动（仅在脉冲触发时计算）
- 节约能耗（0-1信号，稀疏计算）
- 蕴含时间信息
- 神经形态硬件友好

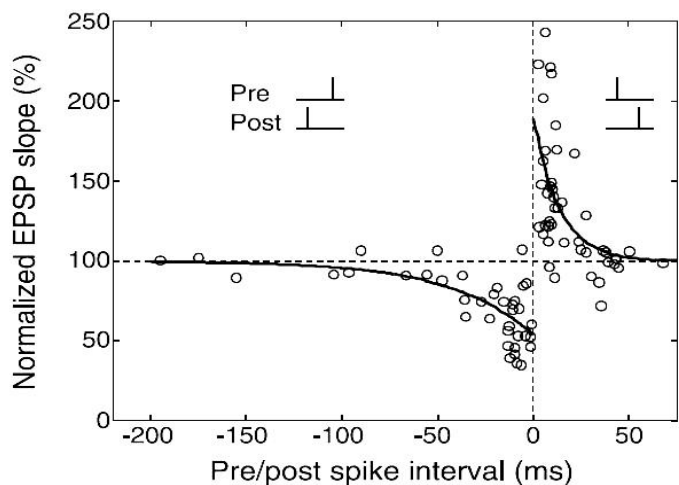


SNN研究意义：

- 实现**高能效、可解释**的类脑智能
- 是**神经形态计算与下一代AI芯片**的核心算法基础
- 在**动态视觉事件相机、脑机接口、低功耗边缘设备**等场景中可以应用

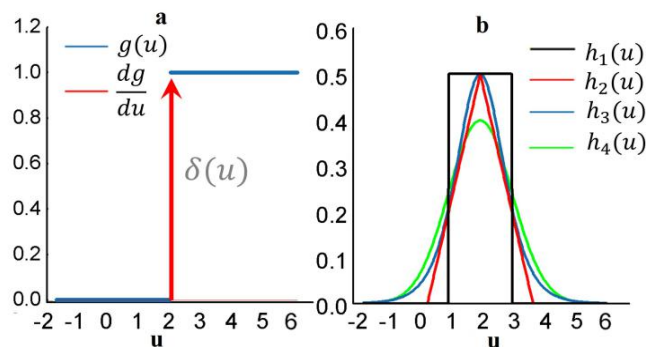
1.2 脉冲神经网络算法研究现状

STDP



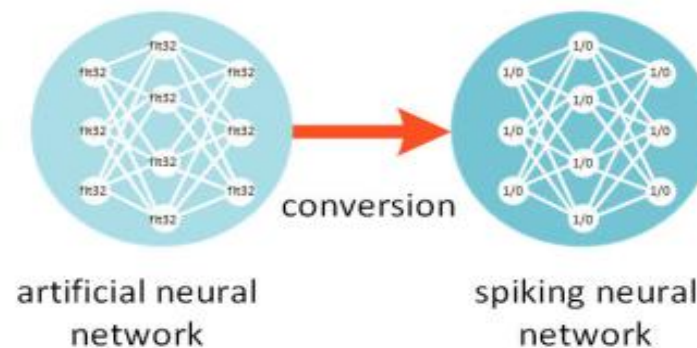
特点：生物启发式，局部可塑性
 缺点：无监督，精度有限

替代梯度下降法



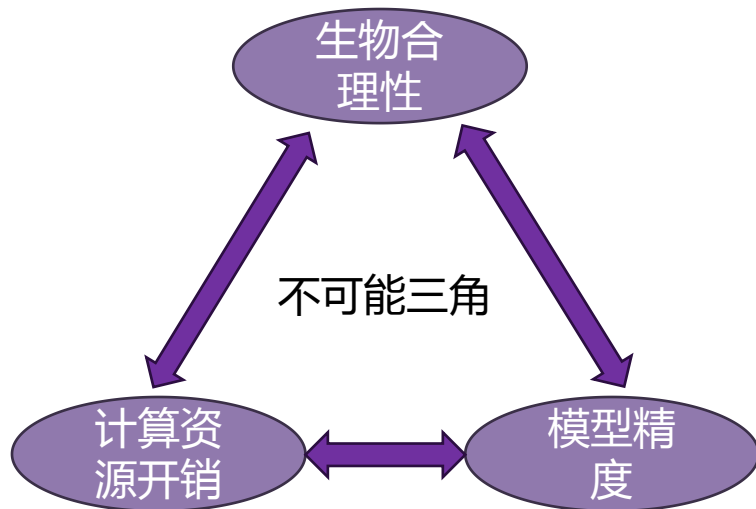
特点：借助代理函数实现反向传播
 缺点：计算内存开销大，生物不合理

ANN-to-SNN方法



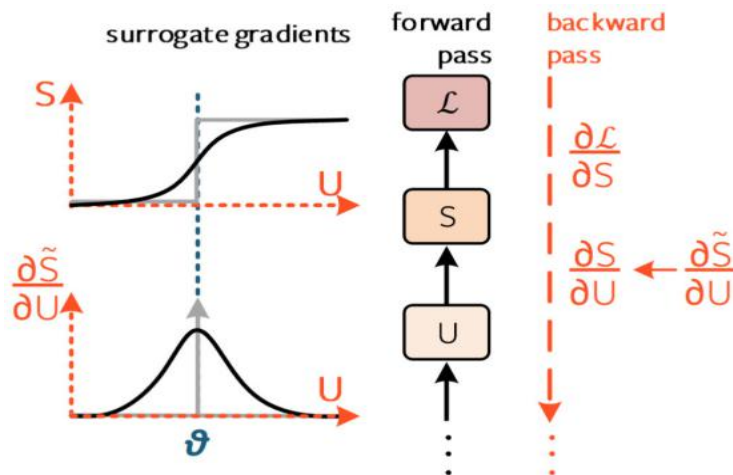
特点：ANN权重迁移，精度高
 缺点：训练推理不一致，生物不合理

1.3 现存问题与挑战



生物合理性、计算能耗和准确率的三元权衡困境，主流SNN算法无法同时满足三个方面。

问题根源：脉冲阶跃函数不可微



膜电压： $u_t = \lambda u_{t-1} + I_t - V_{\text{reset}} \cdot S_{t-1}$

脉冲： $S_t = H(u_t - V_{\text{th}}) = \begin{cases} 1, & \text{if } u_t \geq V_{\text{th}} \text{ (发放脉冲)} \\ 0, & \text{if } u_t < V_{\text{th}} \text{ (不发放)} \end{cases}$

求导： $\frac{dH(u_t - V_{\text{th}})}{du_t} = \delta(u_t - V_{\text{th}})$

(δ 函数仅在阈值处为 ∞ ，其余处为 0)

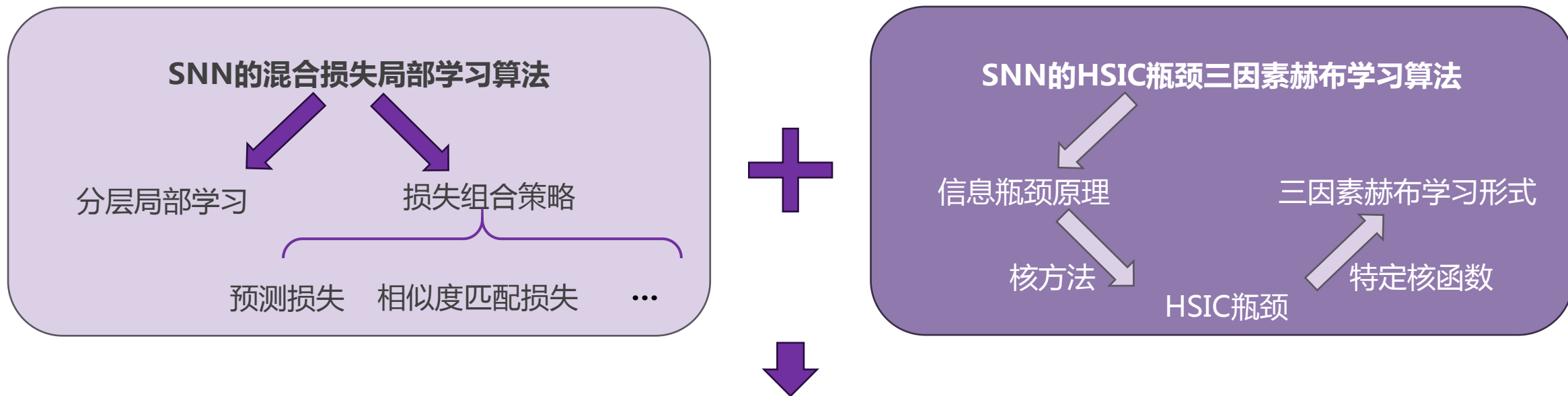
基于BP的SNN训练算法的问题：

- 权重对称问题
 - 更新锁定问题
 - 梯度消失/爆炸
 - BP生物不合理性
- BP自带问题
- “死神经元”问题
 - 代理梯度的近似误差
 - BPTT存算开销大
- BP结合SNN后的新问题

1.4

研究意义

具有生物合理性的脉冲神经网络非 BP 式训练算法



研究意义：

- 提出了一类新型 SNN 训练算法，保持生物合理性的同时，有较好的能效比
- 分层局部训练方式便于并行化，低存算需求对神经形态硬件友好
- 为脉冲神经网络的训练算法提供了新解决方案

第二部分

研究内容

Research Content

研究内容1：基于混合损失局部学习的脉冲神经网络训练算法

研究内容2：基于HSIC瓶颈的脉冲神经网络训练算法

2.1 基于混合损失局部学习的SNN训练算法：研究动机

脉冲神经网络全局BP算法问题

BP算法的缺陷：

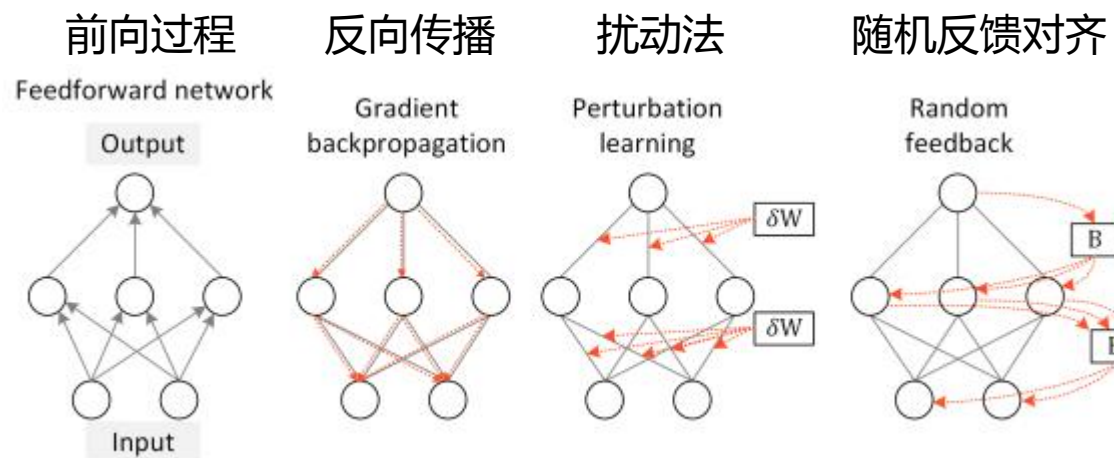
- 权重对称
- 更新锁定
- 梯度消失/爆炸
- 生物不合理性

替代梯度下降法的问题：

- “死神经元”问题
- 代理函数的近似误差累积
- 计算内存开销大



一些SNN非BP式算法的尝试

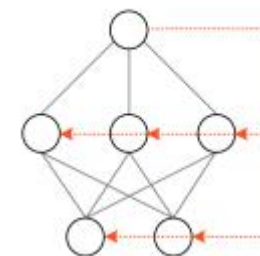


改进

SNN非BP式局部损失学习算法

为 SNN 每一隐藏层单独计算局部损失，各层独立进行局部权值更新，避免全局反向传播 BP 方法的一系列问题。

Local losses



2.1 基于混合损失局部学习的SNN训练算法：局部损失

Layer-wise Local Loss

辅助分类器（层）模块：

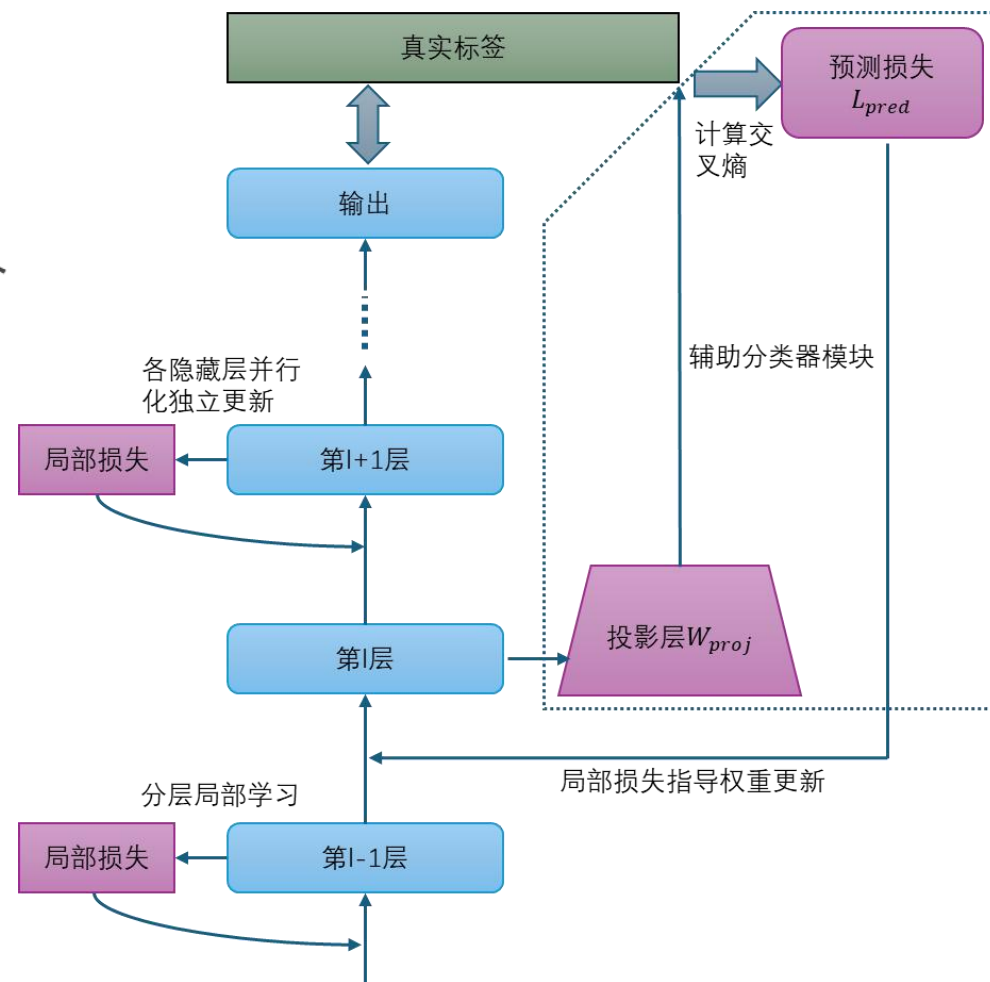
为 SNN 每一隐藏层增加一个辅助分类器（Auxiliary Classifier），对整个网络最终标签进行预测，辅助分类层相当于当前层之后所有层的简化，形式如下：

$$y^*(t) = W_{proj}^l Flatten(s^l[t])$$

计算辅助分类层的预测值和真实标签的交叉熵损失，即为预测损失：

$$\begin{aligned} L_{pred} &= \sum_{t=1}^T CrossEntropy(y, y^*) \\ &= \sum_{t=1}^T CrossEntropy(y, W_{proj}^l Flatten(s^l[t])) \end{aligned}$$

每层的损失只负责当前层的权重更新，局部、独立、并行化。



2.1 基于混合损失局部学习的SNN训练算法：混合局部损失

单一损失类型的局部损失算法，性能无法超越基于BP的方法，怎么办？

损失组合策略：

$$Loss = \alpha L_{pred} + \beta L_{sim} + \gamma L_{contrast} + \zeta L_{recon}$$

$\alpha, \beta, \gamma, \zeta$ 为权重系数，可为0。

对比损失

度量不同样本的脉冲发放率在欧几里得空间的距离关系，得到欧式距离矩阵 D ，对所有正负样本对计算对比损失函数：

$$L_{contrastive} = \sum_{ij} [y_{i,j} D_{i,j}^2 + (1 - y_{i,j}) \max(0, m - D_{i,j})^2]$$

其他类型损失

相似度匹配损失

当前层一个 batch 内样本 H 的压缩特征为 $E_{\theta}(H)$ ，计算其余弦相似度矩阵 $S(E_{\theta}(H))$ ，以及对应标签的 cosine 相似度矩阵，计算两者的 Frobenius 范数：

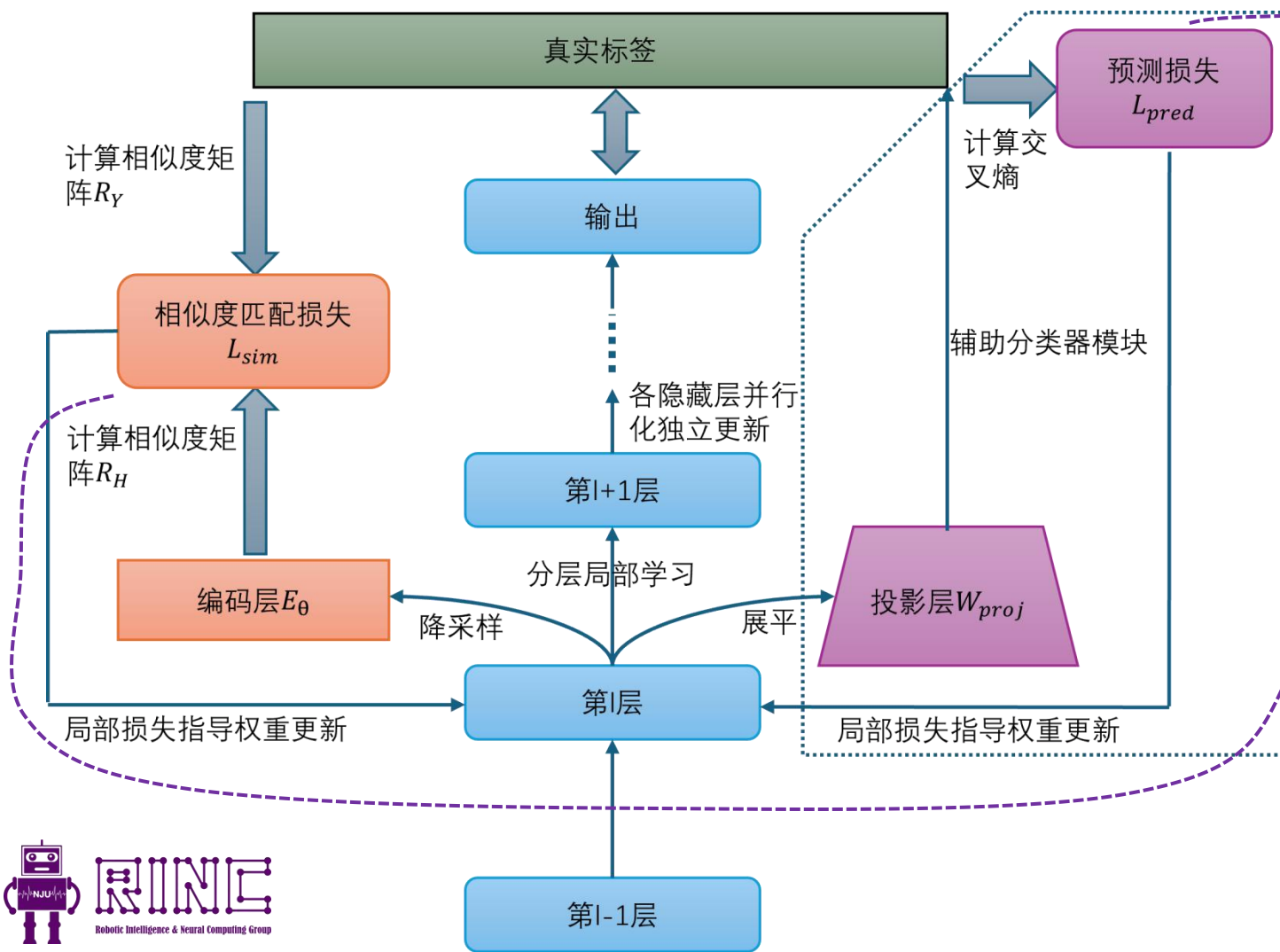
$$L_{similarity} = \sum_{t=1}^T \|S(E_{\theta}(H[t])) - S(Y)\|_F^2$$

编码重构损失

SNN的当前层可视为编码器，附加一个解码器重构当前层的输入，约束SNN每一层尽量多保留有用输入信息，编码重构损失如下：

$$\begin{aligned} L_{reconstruct} &= \|G_l(h_l) - h_{l-1}\|_2^2 \\ &= \|G_l(F_l(h_{l-1})) - h_{l-1}\|_2^2 \end{aligned}$$

2.1 基于混合损失局部学习的SNN训练算法：混合局部损失



①. 预测损失 $L_{predict}$:

衡量SNN当前层对分类任务的判别能力

②. 相似性匹配损失 $L_{similarity}$:

衡量SNN当前层特征在样本空间中的相似性结构和对应标签相似性结构的一致性程度

③. 对比损失 $L_{contrastive}$:

衡量SNN当前层特征正负样本对之间的相对位置关系，使得特征空间在类别边界上更明确。

④. 编码重构损失 $L_{reconstruct}$:

约束SNN每一层尽量多保留原始输入的有效信息

2.1 基于混合损失局部学习的SNN训练算法：消融试验

表 3-5 CIFAR-10 数据集不同损失组合消融实验结果

损失组合名称	损失 1 及系数	损失 2 及系数	损失 3 及系数	测试准确率
预测损失 L_{pred}	$1.0 * L_{pred}$	/	/	87.30%
相似度损失 L_{sim}	$1.0 * L_{sim}$	/	/	87.92%
对比损失 $L_{contrast}$	$1.0 * L_{contrast}$	/	/	78.25%
预测重构损失 $L_{pred-recon}$	$0.5 * L_{pred}$	$0.5 * L_{recon}$	/	81.15%
相似度重构损失 $L_{sim-recon}$	$0.01 * L_{sim}$	$0.99 * L_{recon}$	/	81.54%
预测对比损失 $L_{pred-contrast}$	$0.95 * L_{pred}$	$0.05 * L_{contrast}$	/	78.85%
相似度对比损失 $L_{sim-contrast}$	$0.2 * L_{sim}$	$0.8 * L_{contrast}$	/	60.30%
预测相似度损失 $L_{pred-sim}$	$0.99 * L_{pred}$	$0.01 * L_{sim}$	/	90.77%
三重损失 $L_{pred-sim-contrast}$	$0.95 * L_{pred}$	$0.01 * L_{sim}$	$0.04 * L_{contrast}$	89.91%

- 同实验设置下，预测损失和相似度匹配损失的组合效果最佳。
- 混合损失从不同角度对网络进行优化，能够“优势互补”，相比单一类型损失，有更好的泛化能力和优化效果。

权重系数设定方式：

基于 loss 尺度一致原则，使得在混合损失中不同类型损失分量的尺度大致相当。

2.1 基于混合损失局部学习的SNN训练算法：对比试验

表 3-2 MNIST 数据集实验结果

工作	训练算法	编码方式	网络结构	时间步长	测试准确率
[41]	STDP(unsupervised)	rate code	FCN(784-6400-10)	10	95.0%
[42]	Spatio-temporal BP	rate code	FCN(784-800-10)	10	98.89%
[48]	Global Feedback + STDP	temporal code	FCN(784-800×3-10)	10	98.6%
[31]	STDP-based BP	rate code	CNN	/	97.2%
[34]	DECOLLE(local learning)	/	CNN	10	97.51%
[55]	TSLI(local learning)	direct code	CNN	10	99.35%
[56]	surrogate gradient BP	rate code	CNN	10	99.26%
ours	局部损失 $L_{pred\sim}$	direct code	FCN	10	97.72%
ours	局部损失 $L_{pred\sim}$	direct code	CNN	10	99.35%

- 在MNIST、Fashion-MNIST和Cifar-10数据集上，预测-相似度损失 $L_{pred\sim}$ 相比对照组实验取得最优结果，精度明显提升。
- 本算法兼顾生物合理性、计算内存开销、实验精度等方面。

表 3-3 Fashion-MNIST 数据集实验结果

工作	训练算法	编码方式	网络结构	时间步长	测试准确率
[48]	Global Feedback + STDP	temporal code	FCN	10	89.05%
[32]	Implicit Differentiation	rate code	FCN	5	90.25%
[34]	DECOLLE(local learning)	/	CNN	10	90.75%
[55]	TSLI(local learning)	direct code	CNN	10	92.56%
ours	局部损失 $L_{pred\sim}$	direct code	VGG-like	10	92.11%
ours	局部损失 $L_{pred\sim}$	direct code	VGG-like(2x)	10	93.46%

表 3-4 CIFAR-10 数据集实验结果

工作	训练算法	编码方式	网络结构	时间步长	测试准确率
[42]	Spatio-temporal BP	rate code	VGG8	12	90.53%
[57]	surrogate gradient BP	rate code	VGG9	100	90.45%
[57]	surrogate gradient BP	rate code	ResNet11	100	90.95%
[58]	surrogate gradient BP	rate code	VGG9	25	90.50%
[59]	ANN-to-SNN	temporal code	ResNet20	2048	91.42%
[32]	Implicit Differentiation	rate code	VGG8	30	92.08%
[34]	DECOLLE(local learning)	/	VGG8	10	74.70%
[55]	TSLI(local learning)	direct code	VGG8	10	89.22%
ours	局部损失 $L_{pred\sim}$	direct code	VGG8	10	90.77%
ours	局部损失 $L_{pred\sim}$	direct code	VGG8(2x)	10	91.44%

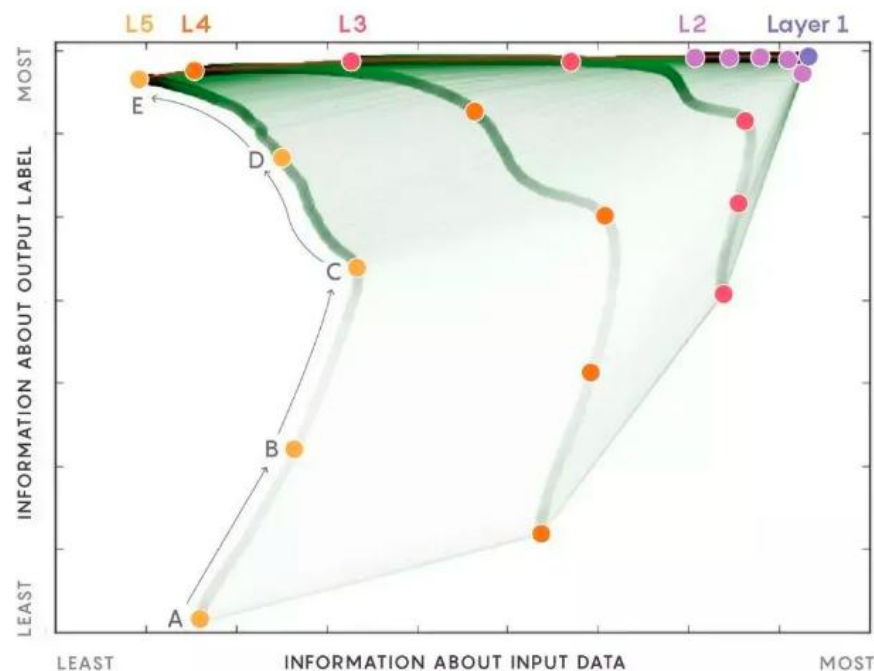
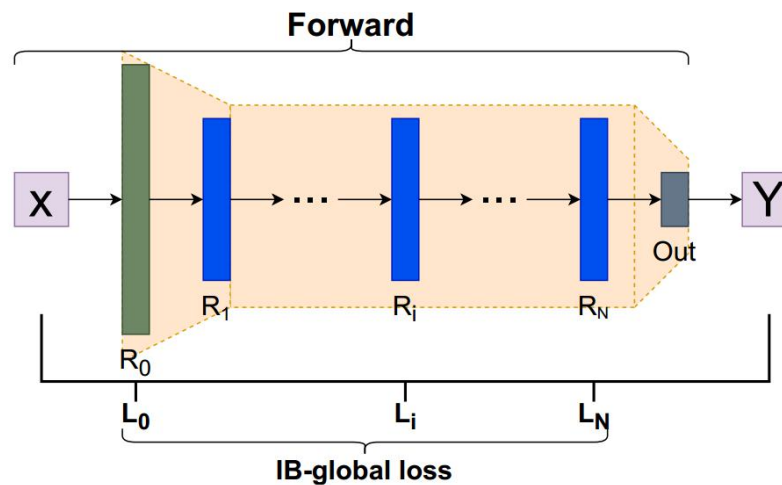
2.2 基于HSIC瓶颈的SNN训练算法：背景介绍

信息瓶颈理论：

信息瓶颈 (Information Bottleneck, IB) 理论是一种理解神经网络学习的**信息论框架**。旨在学习输入 X 的最优表示 T ，并优化以下目标：

$$\mathcal{L}_{IB} = I(X; T) - \beta I(T; Y)$$

- $I(X; T)$ 表示输入 X 和特征 T 的**互信息**
- 最小化 $I(X; T)$ (**压缩**输入冗余信息)
- 最大化 $I(T; Y)$ (**保留**与输出 Y 相关的有用信息)
- 两阶段：**拟合阶段**+**压缩阶段**



2.2 基于HSIC瓶颈的SNN训练算法：信息瓶颈训练SNN的可行性与问题

信息瓶颈训练SNN的可行性：

- 信息瓶颈方法为网络每一层计算信息瓶颈损失 L_{IB} ，是一种**局部损失信号**，可以单独为SNN每一层提供**监督优化信号**，契合SNN的局部学习需求。
- 信息瓶颈为理解神经网络乃至脉冲神经网络的学习过程提供了有别于常规BP式算法的**全新范式**。

信息瓶颈训练SNN的问题：

- 信息瓶颈方法中计算互信息 $I(X; Z)$ 时，需要知晓数据的**真实分布估计**，而真实数据场景难以获取分布。

$$I(X; T) = \int p(x, t) \log \frac{p(x, t)}{p(x)p(t)} dx dt$$

信息瓶颈框架变体

避免分布估计

HSIC瓶颈：
一种**核方法**结合信息瓶颈的算法

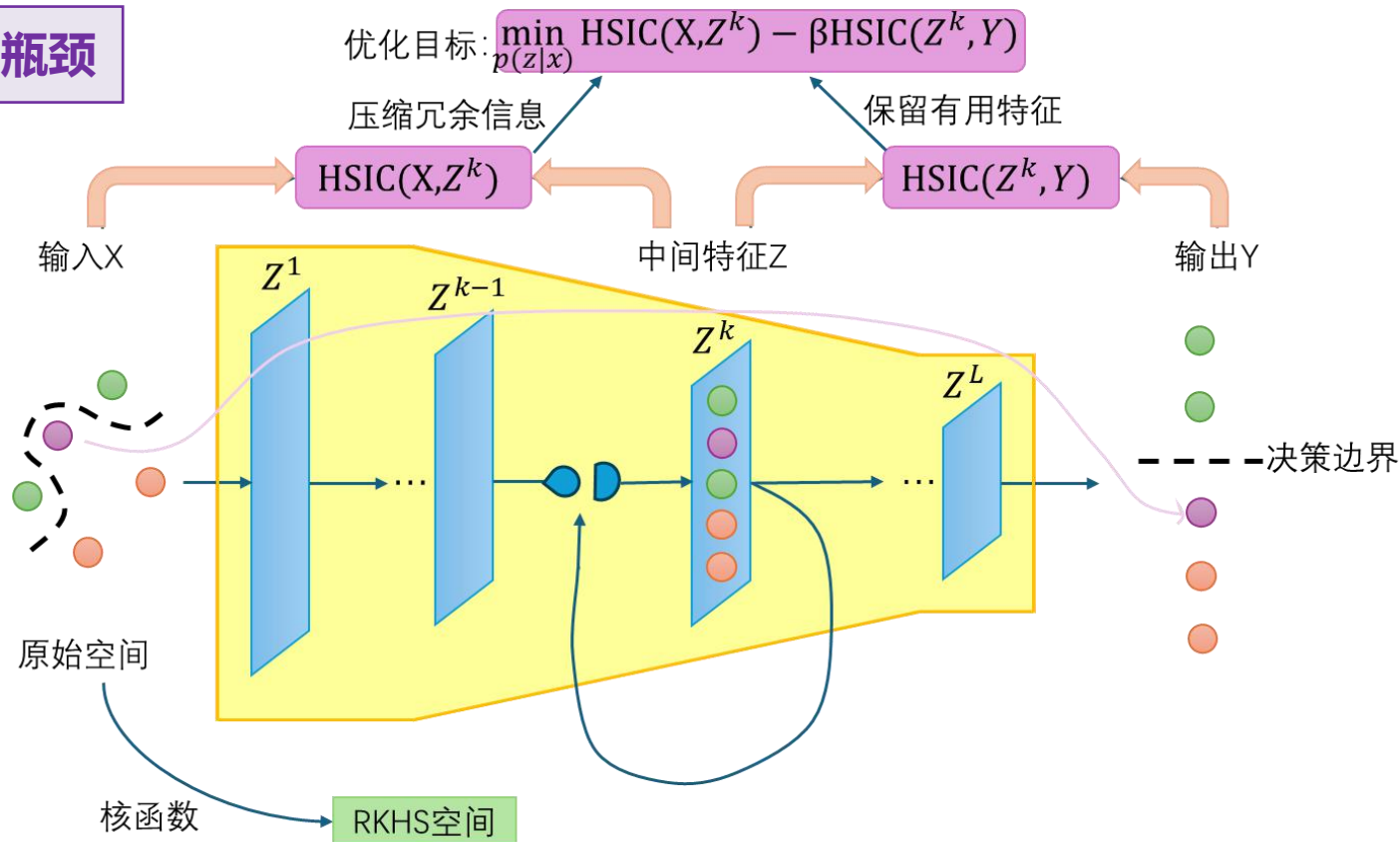
2.2 基于HSIC瓶颈的SNN训练算法：HSIC瓶颈

HSIC瓶颈：

HSIC (希尔伯特-施密特独立性准则) 是一种度量变量**相关性 (依赖性)**的统计量。通过核函数将低维空间数据映射到**高维再生核希尔伯特空间**，显式的分布估计变为隐式的核矩阵计算。

$$\min_{p(\hat{z}|x)} HSIC(X, Z^k) - \beta HSIC(Z^k, Y), k = 1, 2, \dots, L$$

$$HSIC(X, Y) = \frac{1}{m^2} \sum_{i,j} K_1(x_i, x_j) K_2(y_i, y_j) + \frac{1}{m^2} \sum_{i,j} K_1(x_i, x_j) \frac{1}{m^2} \sum_{k,l} K_2(y_k, y_l) - \frac{2}{m^3} \sum_{i,j,l} K_1(x_i, x_j) K_2(y_i, y_l).$$



可将低维空间的**非线性依赖关系**映射到高维空间，用核方法**隐式计算**，避免高维空间的复杂运算。HSIC(X,Z)=0代表独立；HSIC(X,Z)越大代表相关性 (**线性+非线性**) 越明显。

2.2 基于HSIC瓶颈的SNN训练算法：HSIC瓶颈

特定核函数→三因素赫布学习形式：

可选核函数包括 gauss 核，laplace 核，cosine 相似度核等。

当采用高斯核函数时，SNN当前层权重更新呈现显式的赫布学习规则：

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|_2^2}{2\sigma^2}\right).$$

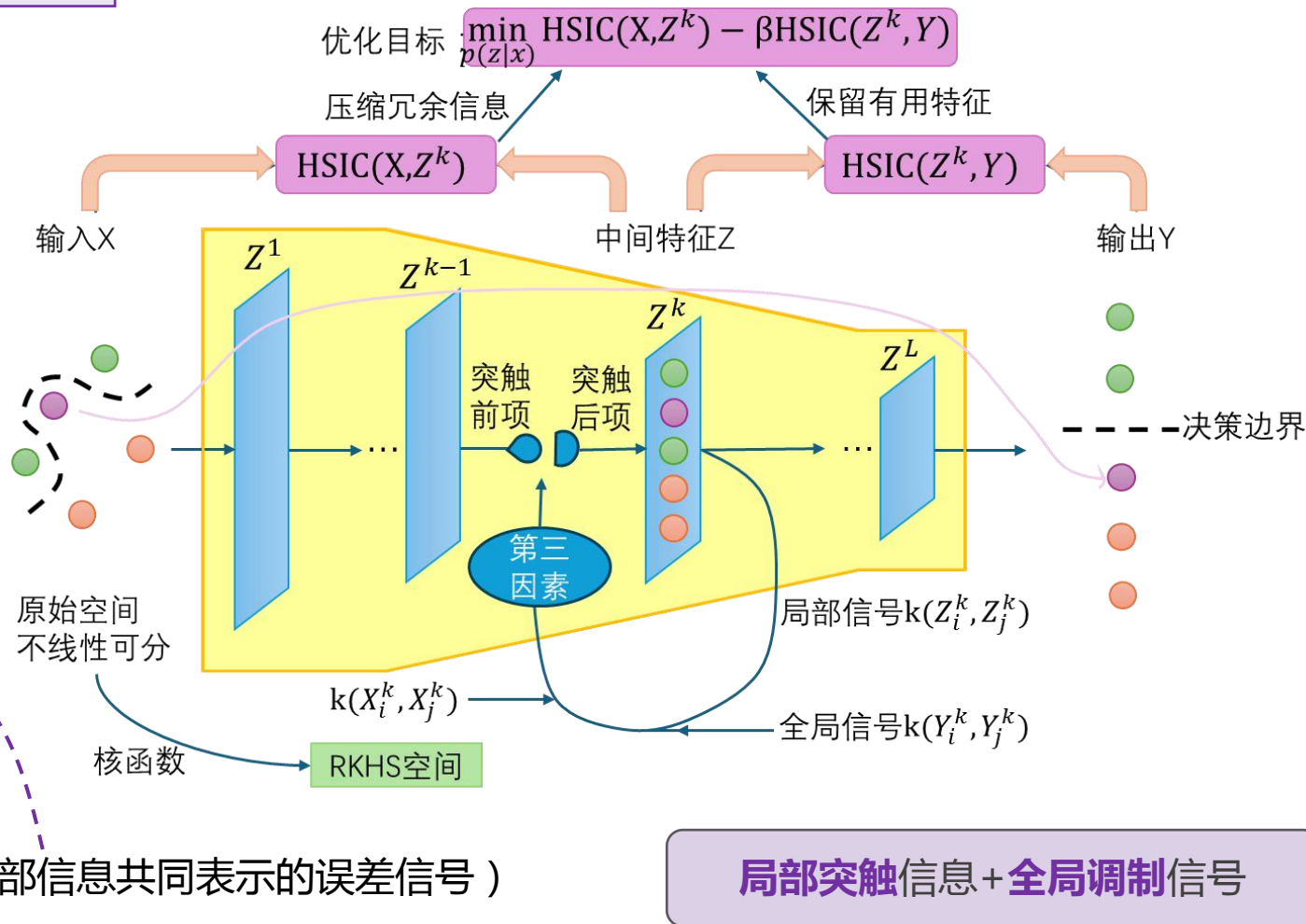
$$\Delta W^k \propto M_{ij}^k (\mathbf{z}_i^k - \mathbf{z}_j^k) (\mathbf{z}_i^{k-1} - \mathbf{z}_j^{k-1})^\top$$

$$M_{ij}^k = -\frac{1}{\sigma^2} \left(\gamma \dot{k}(\mathbf{y}_i, \mathbf{y}_j) - 2 \dot{k}(\mathbf{z}_i^k, \mathbf{z}_j^k) \right) k(\mathbf{z}_i^k, \mathbf{z}_j^k).$$

突触后项

突触前项

第三因素（全局和局部信息共同表示的误差信号）



2.2 基于HSIC瓶颈的SNN训练算法：对比实验

对比实验结果：

基于HSIC瓶颈的SNN局部学习算法，**实验精度上超越**了其他具有生物可信度的SNN算法（如STDP-BP，反馈对齐FA，原始信息瓶颈IB等），接近基于全局BP的SNN算法（替代梯度下降+BPTT）。

可诱导出三因素赫布学习形式的高斯核函数性能最佳。

表 4-4 不同核函数在 Fashion-MNIST 数据集上的效果对比

核函数	表达式	准确率
高斯核	$\exp(-\frac{\ x_1-x_2\ _2^2}{2\sigma^2})$	89.8%
余弦相似度核	$\frac{x_1^T x_2}{\ x_1\ _2 \ x_2\ _2}$	88.7%
拉普拉斯核	$\exp(-\frac{\ x_1-x_2\ _2}{\sigma})$	88.8%
sigmoid 核	$\tanh(\alpha x_1^T x_2 + c)$	87.5%

表 4-1 MNIST 数据集实验结果

工作	训练算法	编码方式	网络结构	时间步长	测试准确率
[41]	STDP(unsupervised)	rate code	FCN(784-6400-10)	10	95.0%
[48]	Global Feedback + STDP	temporal code	FCN(784-800×3-10)	10	98.6%
[42]	Spatio-temporal BP	rate code	FCN(784-800-10)	10	98.89%
[42]	Spatio-temporal BP	rate code	CNN	10	99.42%
[60]	Feedback Alignment	rate code	CNN	10	99.01%
[31]	STDP-based BP	rate code	CNN	/	97.2%
[56]	surrogate gradient BP	rate code	CNN	10	99.26%
[34]	DECOLLE(non-BP)	/	CNN	10	97.51%
[54]	IB(non-BP)	rate code	CNN	3	98.96%
ours	HSIC bottleneck(non-BP)	direct code	CNN	10	99.2%

表 4-2 Fashion-MNIST 数据集实验结果

工作	训练算法	编码方式	网络结构	时间步长	测试准确率
[48]	Global Feedback + STDP	temporal code	FCN	10	89.05%
[32]	Implicit Differentiation	rate code	FCN	5	90.25%
[61]	IB + surrogate gradient BP	temporal code	VGG9	10	90.17%
[54]	IB(non-BP)	rate code	VGG8	3	87.92%
ours	HSIC(non-BP)	direct code	VGG8	10	89.8%

2.2 基于HSIC瓶颈的SNN训练算法：实验分析

基于HSIC瓶颈的SNN局部学习算法的优点：

- 作为具有生物合理性的**局部类赫布学习**算法的同时，保持了一定的模型精度竞争力。
- 可并行化计算更新，计算图简单，**节约存储**资源。
- 可度量变量间**非线性依赖性**，可处理复杂关系数据；对**噪声**输入扰动有一定**鲁棒性**。

表 4-5 加噪声干扰的 MNIST 和 Fashion-MNIST 数据集实验结果

数据集	噪声类型	训练算法	准确率下降
MNIST	高斯噪声	全局替代梯度 BP	7.0%
		HSIC 瓶颈赫布学习	4.4%
	椒盐噪声	全局替代梯度 BP	4.1%
		HSIC 瓶颈赫布学习	2.4%
Fashion-MNIST	高斯噪声	全局替代梯度 BP	12.3%
		HSIC 瓶颈赫布学习	8.6%
	椒盐噪声	全局替代梯度 BP	8.5%
		HSIC 瓶颈赫布学习	6.4%

表 4-6 SNN 非 BP 式训练算法对比

方法	局部算法	非权重对称	非更新锁定	异步可并行化	有竞争力的精度
surrogate-BP	✗	✗	✗	✗	✓
FA	✗	✓	✗	✗	✗
DFA	✓	✓	✗	✗	✗
TP	✓	✓	✗	✗	✓
Eprop	✓	✓	✗	✗	✓
STDP	✓	✓	✓	✓	✗
Our Work 1	✓	✓	✓	✓	✓
Our Work 2	✓	✓	✓	✓	✓

第三部分

研究生期间工作成果

Work Product

相关成果列举

专利

- 徐百乐, **胡嘉骏**, 申富饶。《一种具有生物合理性的分层局部混合损失脉冲神经网络训练算法》。专利申请号: 202510600130.7

项目

- 科技部重大项目“基于神经可塑性的脉冲网络高效学习机制与类脑智能系统”(参与课题年限 2022 年 9 月-2025 年 6 月), 负责脉冲神经网络算法相关研究。

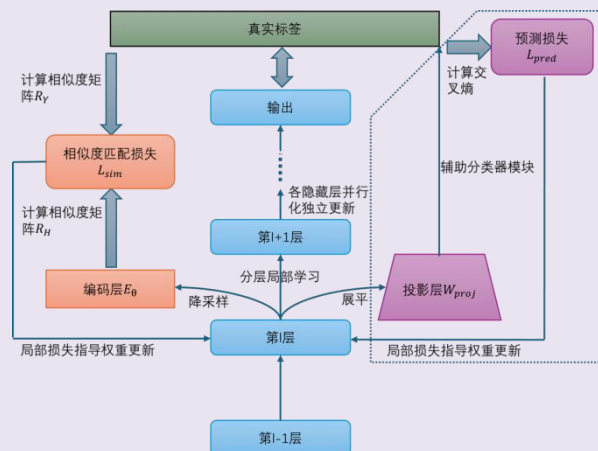
第四部分

总结

Summary

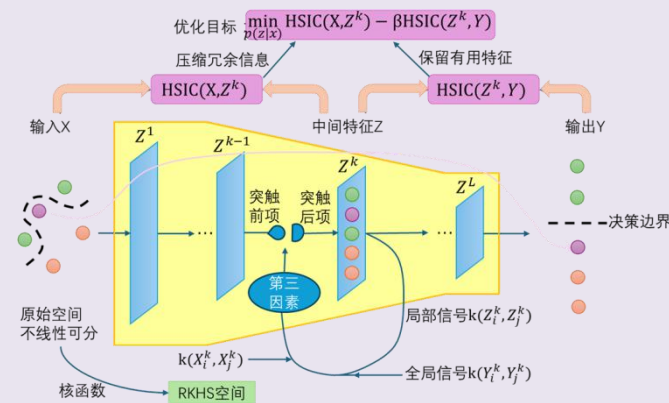
全文总结

基于混合损失局部学习的脉冲神经网络训练算法



- 提出了一种**分层局部损失**算法，并用**混合损失策略**提升模型精度，超越了基于全局BP的SNN算法。
- 分层局部学习有利于**并行化计算**，无需存储大量中间状态，兼顾生物合理性，存算开销和精度。

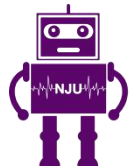
基于HSIC瓶颈的脉冲神经网络训练算法



- 提出了一种**HSIC信息瓶颈**SNN新学习范式，可推导出赫布学习形式的规则，有**生物可解释性**。
- 可并行化计算更新，节约显存资源。
- 可处理**非线性依赖关系**，对噪声数据有**鲁棒性**。



南京大學
NANJING UNIVERSITY



RINC
Robotic Intelligence & Neural Computing Group

感谢各位老师批评指正

RINC GROUP PRESENTATION TEMPLATE

答辩人：胡嘉骏 502022370015

导师：申富饶 教授

日期：2025年5月6日

誠樸雄偉 勵學敦行