上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

SJTU

类脑智能大作业汇报

刘翰文

2024年12月31日

饮 水 思 源 • 爱 国 荣 校

24组

目 录

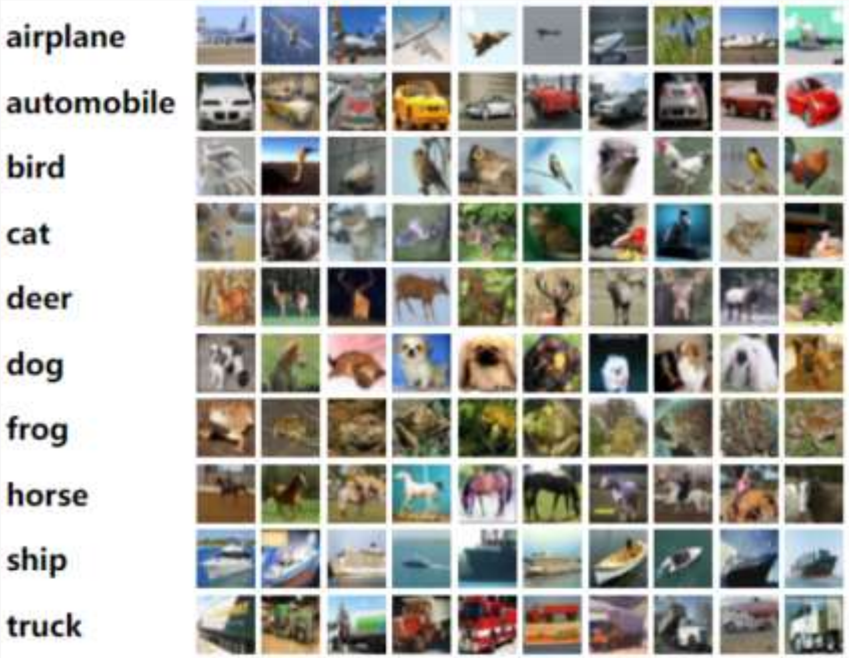
基础部分

01

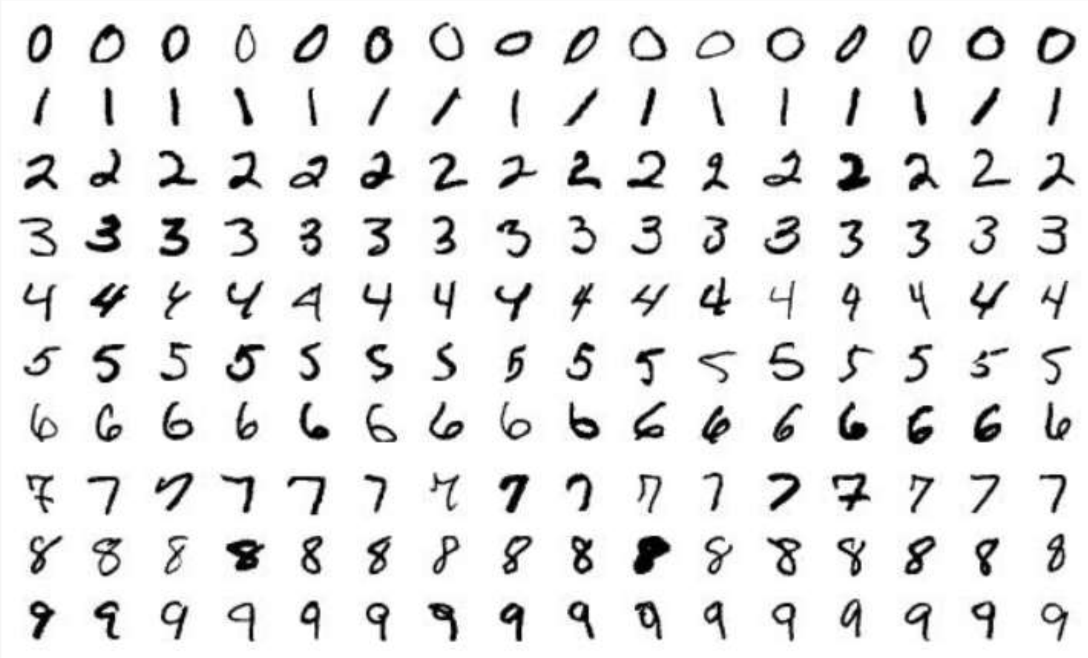
使用Spiking-VGG、ResNet对CIFAR-10分类

数据集

CIFAR-10, MNIST

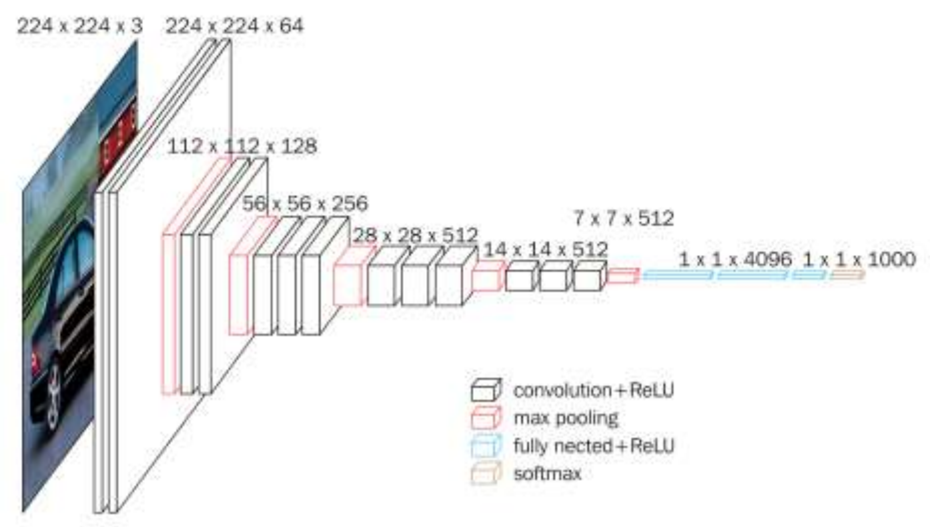


CIFAR-10

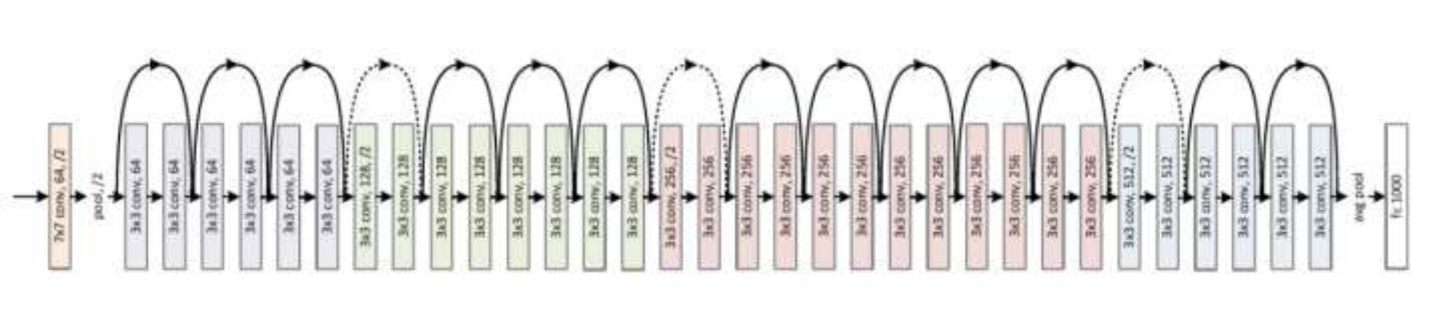


MNIST

VGG16, ResNet34

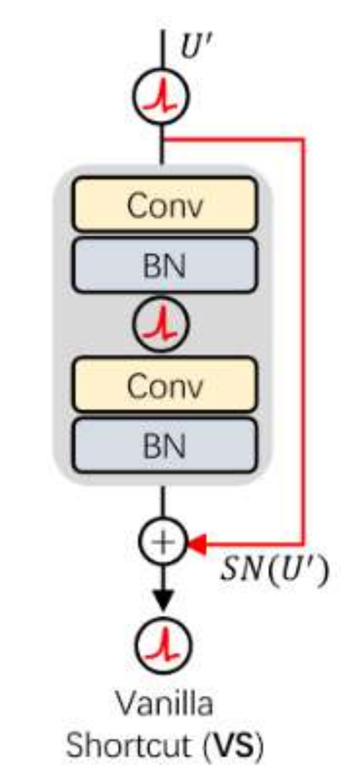


VGG16(ANN)



ResNet34(ANN)

- Spiking VGG16: 90.18% (CIFAR-10)
- Spiking ResNet34: 10.00%! (CIFAR-10)



In Vanilla Shortcut(VS):

$$O'_A = ReLU(F(ReLU(F(U'_A))) + U'_A)$$
$$O'_S = SN(F(SN(F(U'_S)) + U'_S)$$

当 $F(ReLU(F(U'_A))) \equiv 0$，$F(SN(F(U'_S))) \equiv 0$ 时，$O'_A = ReLU(U'_A)$，$O'_S = SN(U'_S)$。

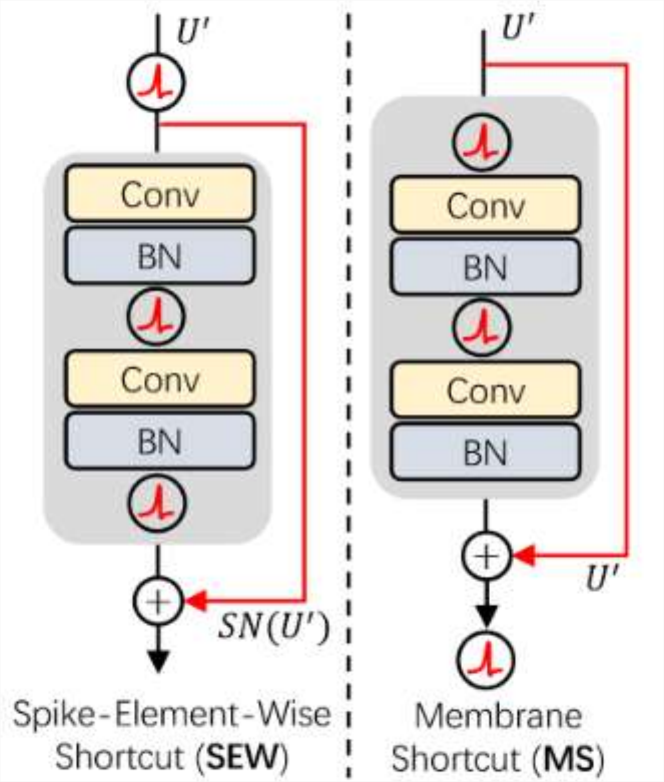
残差连接要求实现恒等映射

$O'_A = ReLU(U'_A) = U'_A$，
$O'_S = SN(U'_S) \neq U'_S$

无法实现恒等映射

Any other shortcut?

- Spike-Element-Wise Shortcut (SEW)
- Membrane Shortcut (MS)



$$O' = SN(F(SN(F(U')))) + U'$$
$$O' = F(SN(F(SN(U')))) + U'$$

$SN\left(F\left(SN(F(U'))\right)\right) \equiv 0, F\left(SN\left(F(SN(U'))\right)\right) \equiv 0$时，$O' = U'$

实现了恒等映射

结果

On CIFAR-10

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
VGG-16	14.79	317.0	64.70	6	**90.18**
SEW-ResNet-34	0.47	72.5	14.80	6	89.03
MS-ResNet-34	0.47	72.5	**8.30**	6	82.39

On MNIST

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
VGG-16	14.79	315.8	64.63	6	99.03
SEW-ResNet-34	0.46	72.2	14.97	6	**99.18**
MS-ResNet-34	0.46	72.2	**8.61**	6	99.13

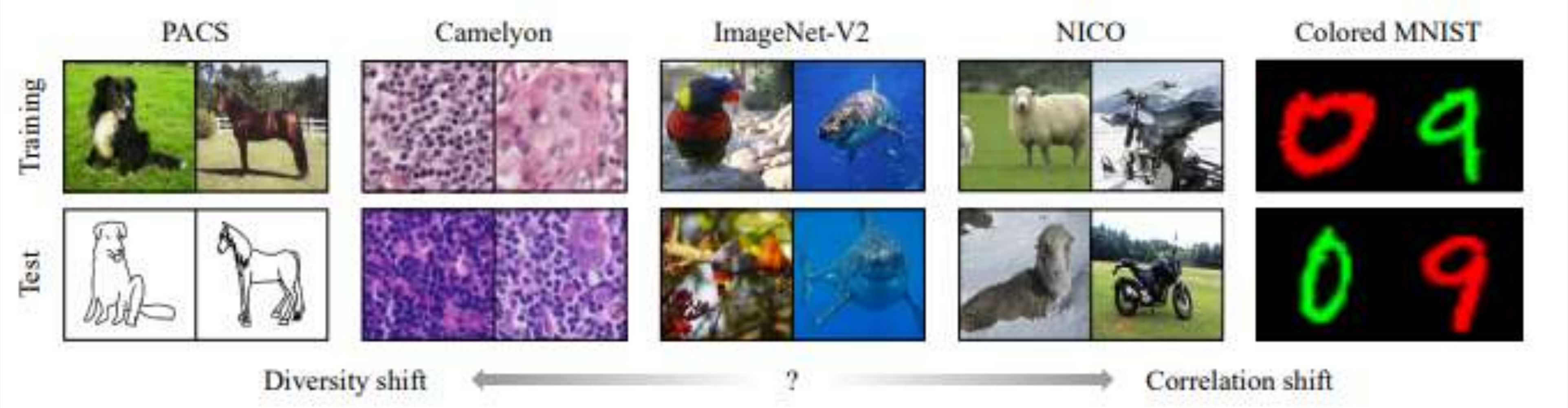
02 中级部分

基于脉冲序列相似性的分布外泛化算法

24组

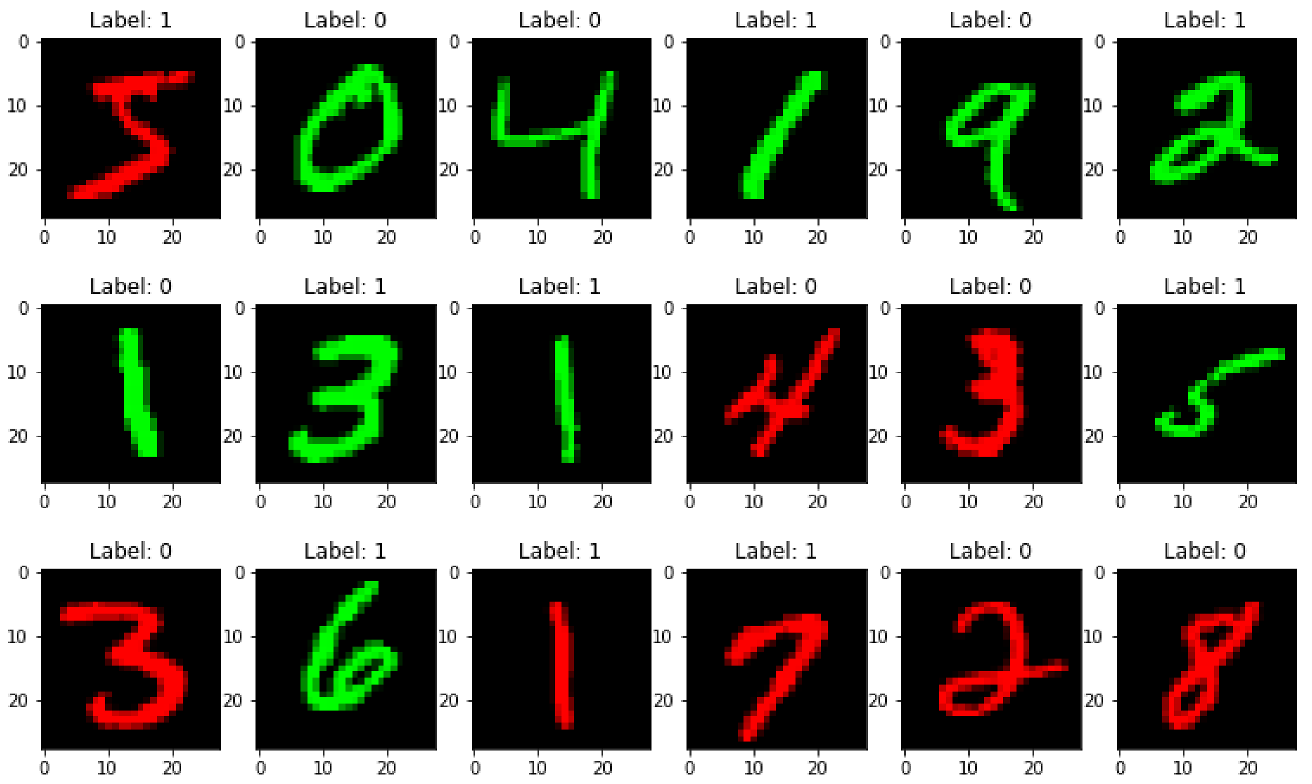
分布外泛化

- 模型在与训练集环境不同的测试集环境中的表现



Colored MNIST

Train1: 10%翻转， Train2: 20%翻转， Test: 90%翻转



脉冲序列相似性

余弦序列相似度

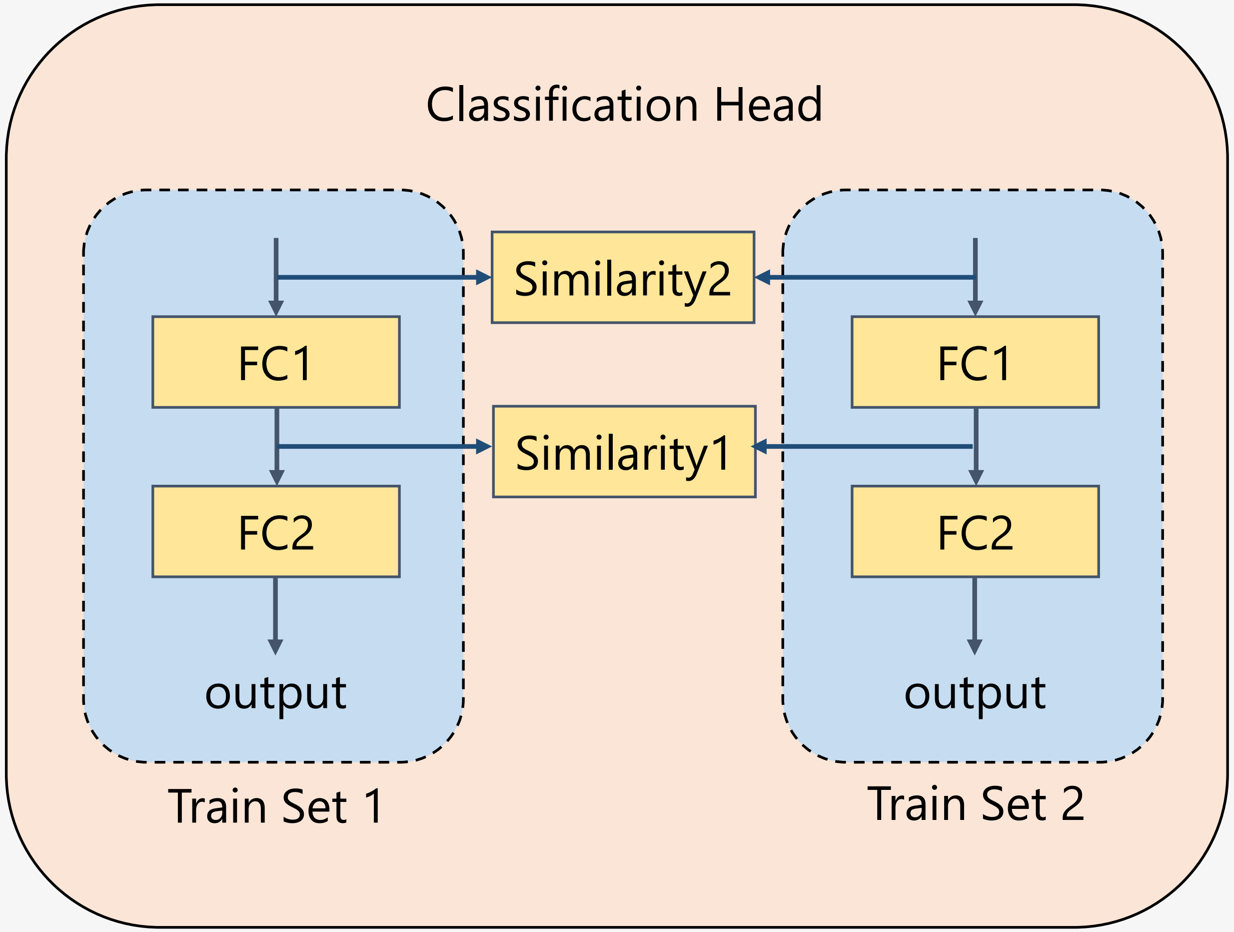
$$Similarity(A,B) = \frac{A \cdot B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n}(A_i \times B_i)}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

训练损失

$$\mathcal{L}_{total} = \mathcal{L}_{crossentropy} + \lambda \cdot \mathcal{L}_{similarity}$$

$$\mathcal{L}_{similarity} = 1 - Similarity(O_1, O_2)$$

λ设置：从0开始随epoch线性增长



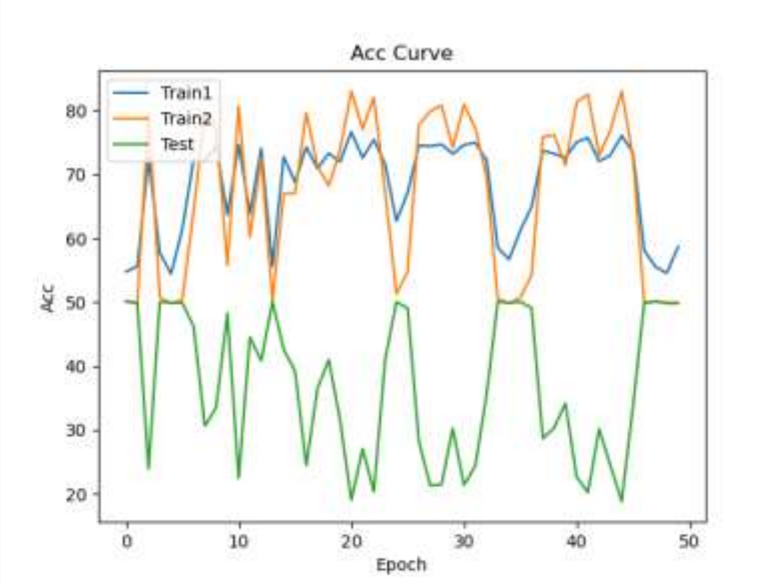
结果

- 在Colored MNIST上训练50个epochs
- 选择在训练集上平均准确率超过70%时的最高测试准确率

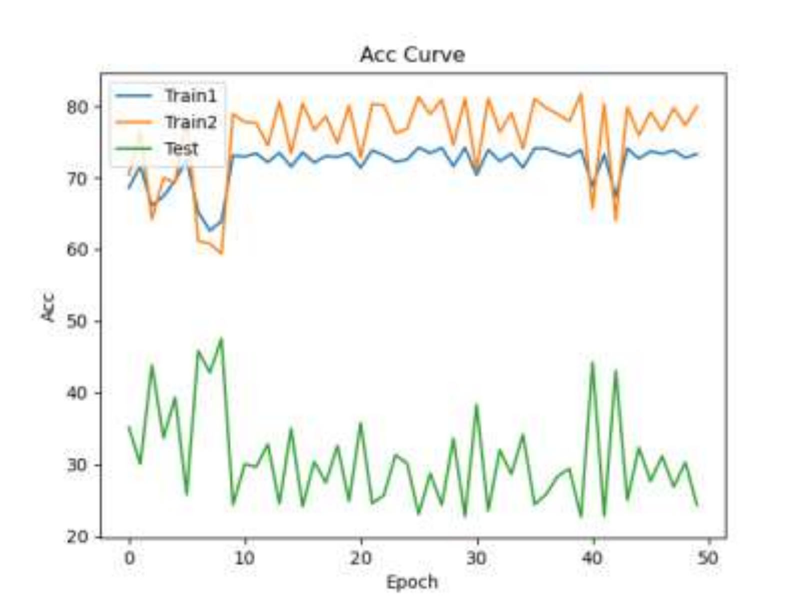
	Train Set 1(%)	Train Set 2(%)	Test Set(%)
None	89.1	80.1	10.2
IRM	73.2	68.1	41.2
Similarity1	75.8	65.0	**42.1**
Similarity2	73.4	73.2	32.5

结果

- IRM: 不稳定，经常出现失败的情况，且准确率不高
- Similarity: 较稳定，但准确率同样不高



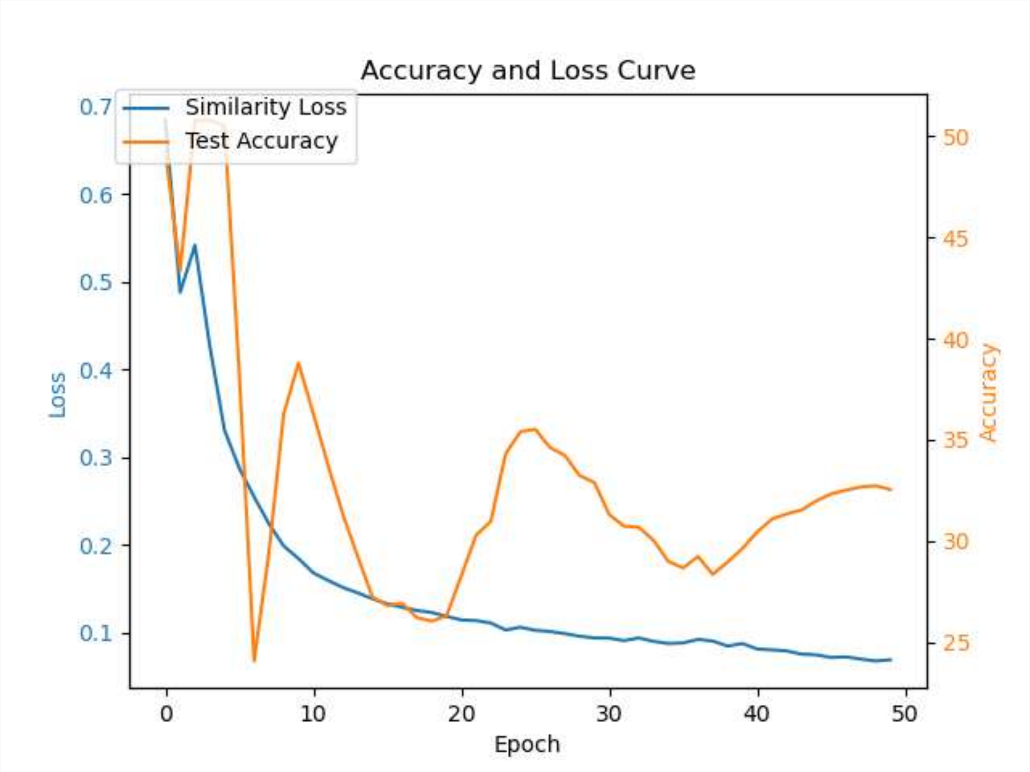
IRM训练过程



Similarity训练过程

分析

跟踪相似性损失和测试集准确率



训练环境差异的影响

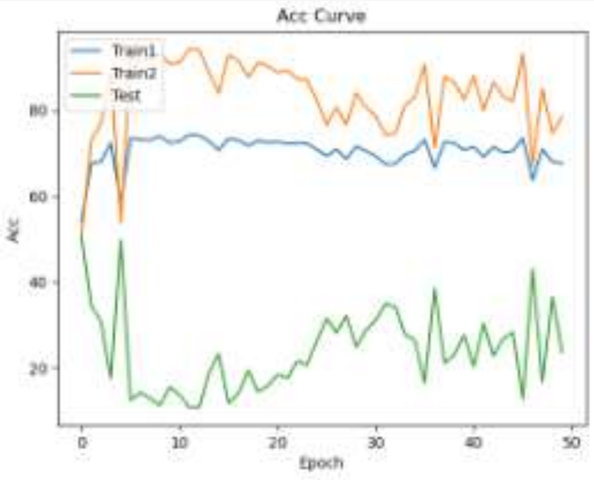
修改两个训练环境翻转的比例

	Train Set 1 Flip(%)	Train Set 2 Flip(%)	Test Set Flip(%)
Colored MNIST	10	20	90
Colored MNIST-1	5	25	90
Colored MNIST-2	2.5	27.5	90
Colored MNIST-3	0	30	90

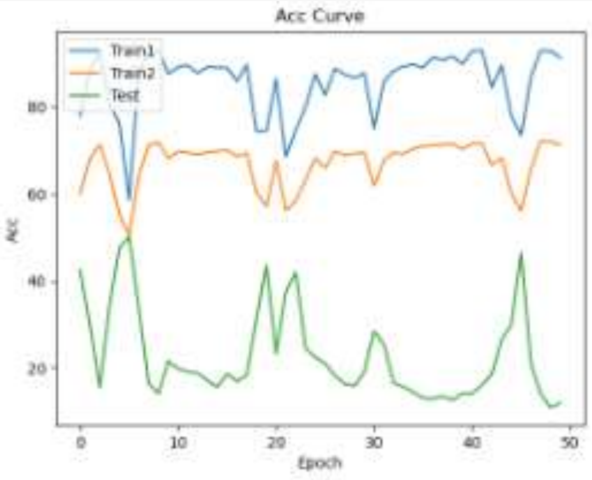
使用相似性算法的结果

	Train Set 1(%)	Train Set 2(%)	Test Set(%)
Colored MNIST	75.8	65.0	42.1
Colored MNIST-1	75.4	67.9	35.1
Colored MNIST-2	87.2	67.9	23.7
Colored MNIST-3	77.4	64.5	33.6

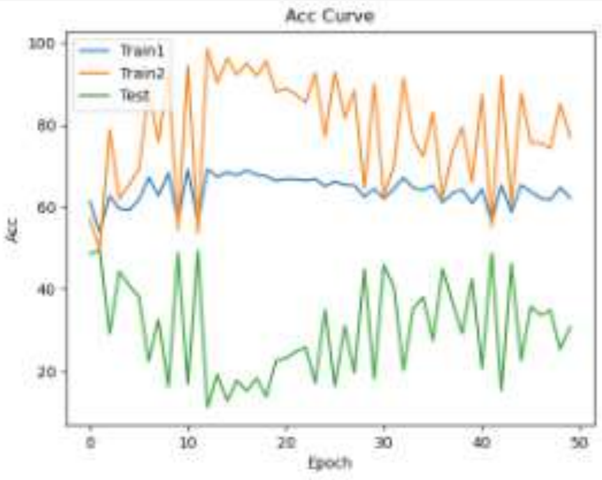
训练环境差异的影响



(d) 5% 和 25% 结果



(e) 2.5% 和 27.5% 结果



(f) 0% 和 30% 结果

03

高级部分

Spiking GhostNet及改进

Ghost Net架构

Input	Operator	#exp	#out	SE	Stride
$224^2 \times 3$	Conv2d 3×3	-	16	-	2
$112^2 \times 16$	G-bneck	16	16	-	1
$112^2 \times 16$	G-bneck	48	24	-	2
$56^2 \times 24$	G-bneck	72	24	-	1
$56^2 \times 24$	G-bneck	72	40	1	2
$28^2 \times 40$	G-bneck	120	40	1	1
$28^2 \times 40$	G-bneck	240	80	-	2
$14^2 \times 80$	G-bneck	200	80	-	1
$14^2 \times 80$	G-bneck	184	80	-	1
$14^2 \times 80$	G-bneck	184	80	-	1
$14^2 \times 80$	G-bneck	480	112	1	1
$14^2 \times 112$	G-bneck	672	112	1	1
$14^2 \times 112$	G-bneck	672	160	1	2
$7^2 \times 160$	G-bneck	960	160	-	1
$7^2 \times 160$	G-bneck	960	160	1	1
$7^2 \times 160$	G-bneck	960	160	-	1
$7^2 \times 160$	G-bneck	960	160	1	1
$7^2 \times 160$	Conv2d 1×1	-	960	-	1
$7^2 \times 960$	AvgPool 7×7	-	-	-	-
$1^2 \times 960$	Conv2d 1×1	-	1280	-	1
$1^2 \times 1280$	FC	-	1000	-	-

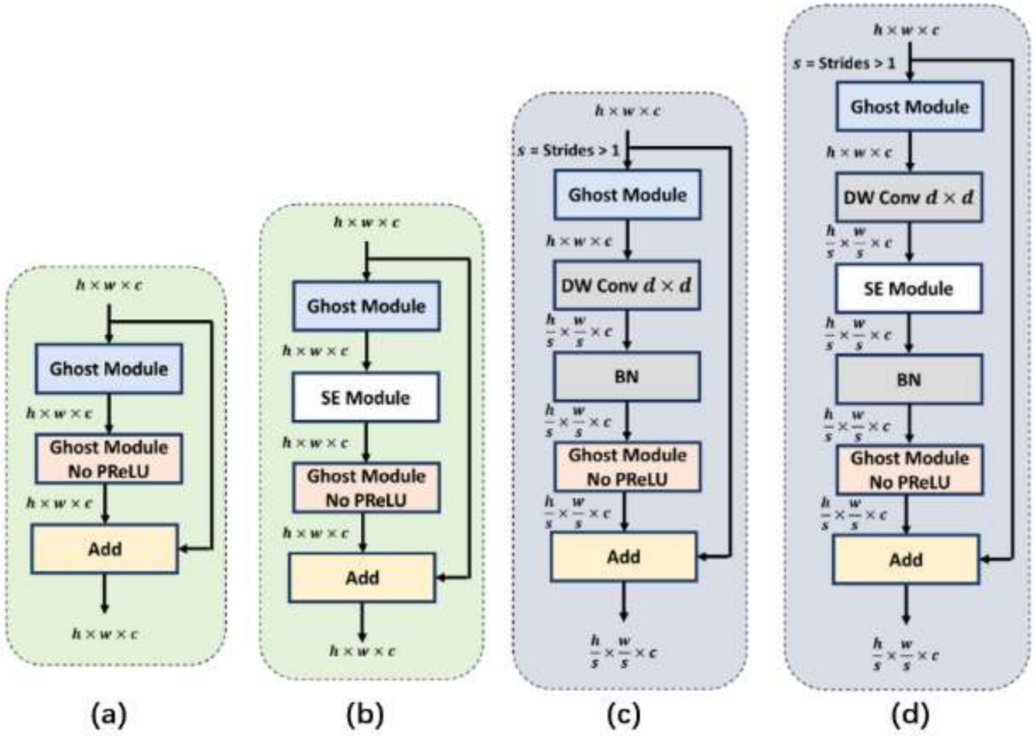
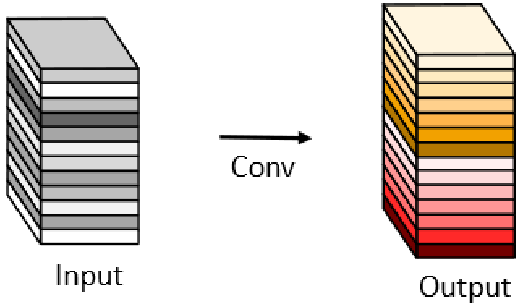
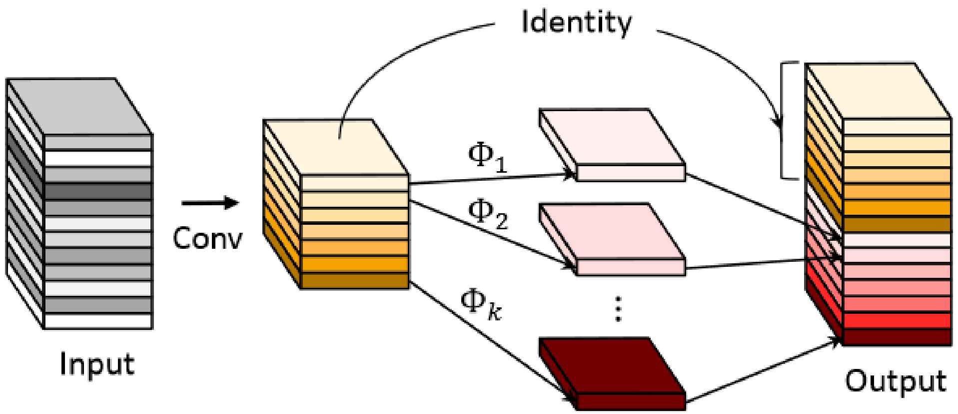


图 2: (a) 无下采样且不带 SE 模块的 Ghost bottleneck (b) 无下采样但带 SE 模块的 Ghost bottleneck (c) 带下采样和 SE 模块的 Ghost bottleneck (d) 带下采样但不带 SE 模块的 Ghost bottleneck

- Ghost Module
- 目的：低计算量同时减轻特征图冗余
- 1. Primary Conv: 轻量卷积 n/m特征图
- 2. Cheap Operation: m-n/m 特征图
- 3. 特征图拼接: m 特征图



(a) The convolutional layer.



(b) The Ghost module.

- Ghost Net：低计算量
- Spiking Ghost Net：低计算量，低能耗
- 结果：CIFAR-10

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
GhostNet-SNN	3.9	6.4	1.04	6	80.12

- 对比基础部分的网络

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
VGG-16	14.79	317.0	64.70	6	**90.18**
SEW-ResNet-34	0.47	72.5	14.80	6	89.03
MS-ResNet-34	0.47	72.5	**8.30**	6	82.39

- 改进一：DFC注意力(Decoupled Fully Connected Attention)
- 特点：简单高效易部署、低计算量

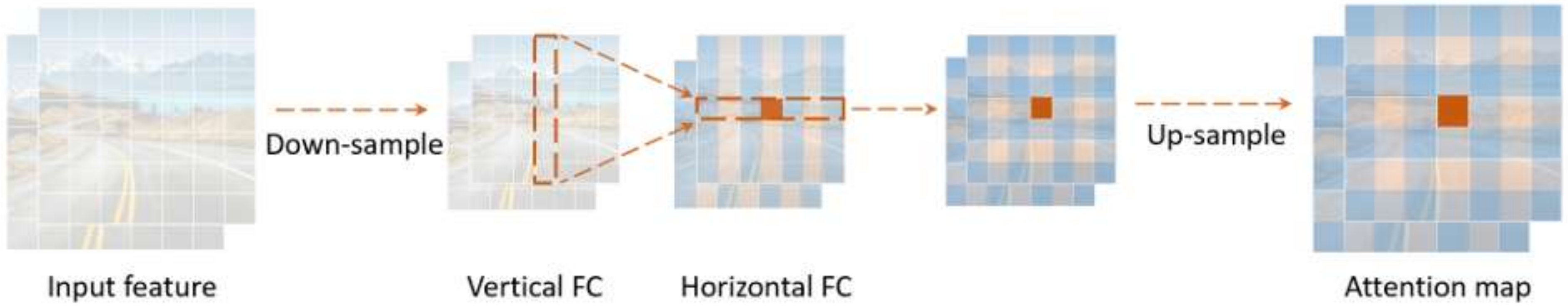
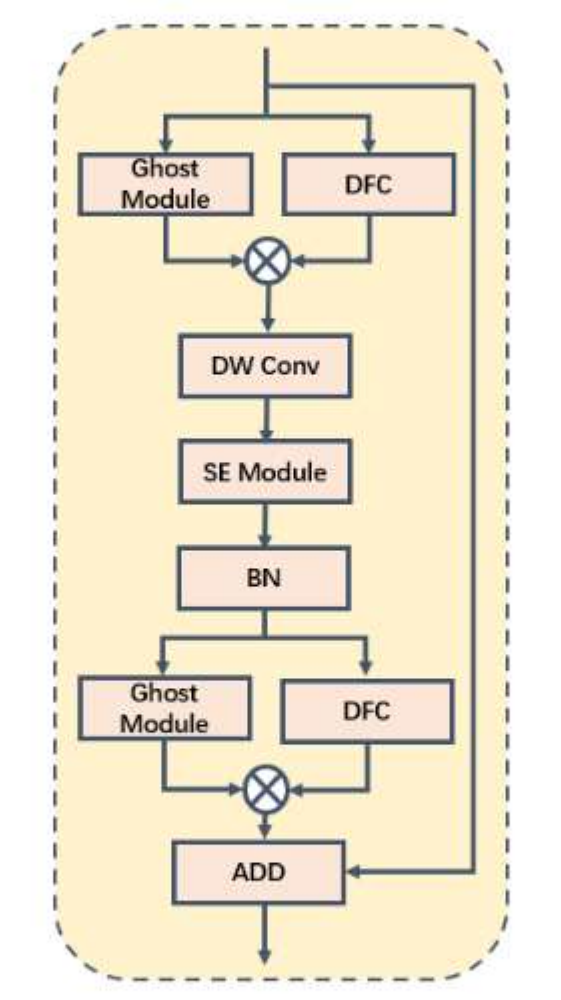


Figure 3: The information flow of DFC attention. The horizontal and vertical FC layers capture the long-range information along the two directions, respectively.

- 加入DFC后的Ghost bottleneck：
- 结果：

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
GhostNet-SNN	3.9	6.4	1.04	6	80.12
GhostNet-DFC	4.9	9.7	1.53	6	80.42



改进

改进二：Channel MLP

参考Meta Transformer Block

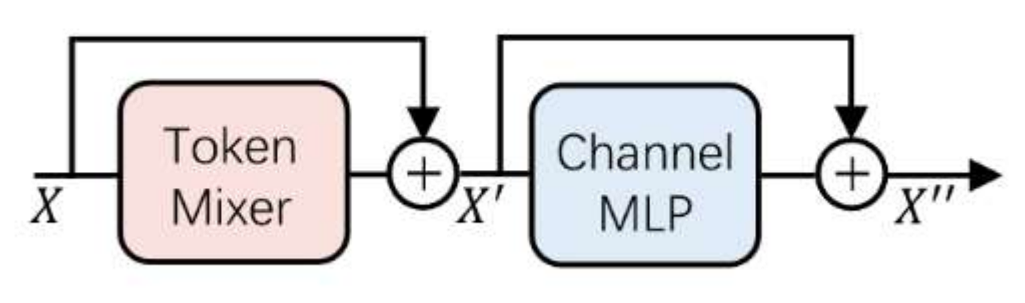
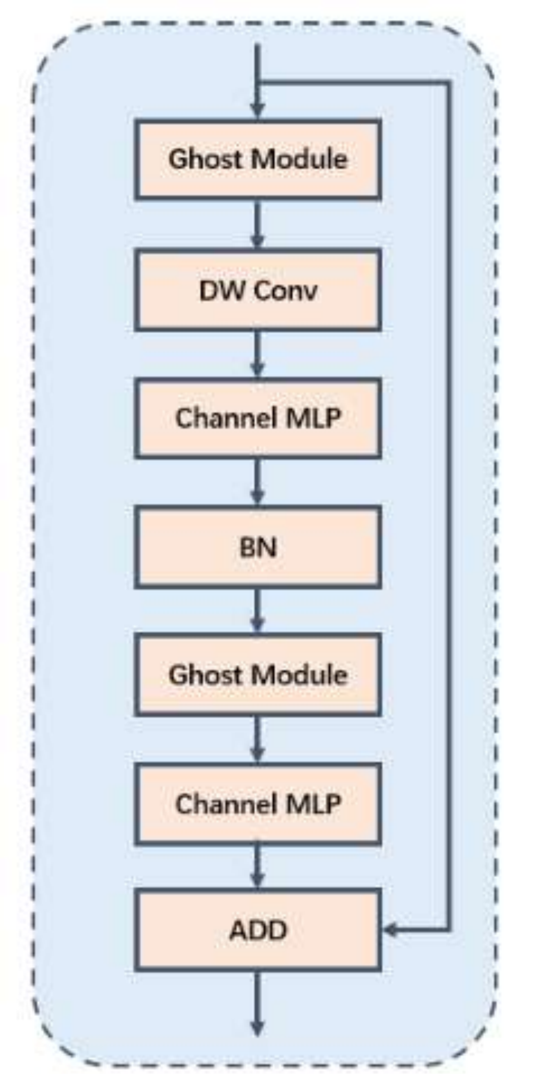


图 8: Meta Transformer Block 架构



Token Mixer：Ghost Module

Channel MLP：+Channel MLP

残差：使用bottleneck的残差

整体结果：

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
GhostNet-SNN	3.9	6.4	1.04	6	80.12
GhostNet-DFC	4.9	9.7	1.53	6	80.42
GhostNet-ChannelMLP	4.7	7.1	**0.96**	6	**81.30**

表 4: GhostNet 在 CIFAR-10 上的实验结果

网络架构	模型参数 (M)	FLOPs(M)	能耗 (μJ)	时间步长	准确率 (%)
VGG-16	14.79	317.0	64.70	6	**90.18**
SEW-ResNet-34	0.47	72.5	14.80	6	89.03
MS-ResNet-34	0.47	72.5	**8.30**	6	82.39

表 1: 网络在 CIFAR-10 上的实验结果

Thank You

饮 水 思 源 爱 国 荣 校

24组