

CSWin-Transformer 模型自验报告

林丽森 finder4alex@163.com

1. 模型简介

1.1 网络模型结构简介

CSWin Transformer 是 Swin 原班人马在2021年7月提出的网络结构，网络结构见下图。

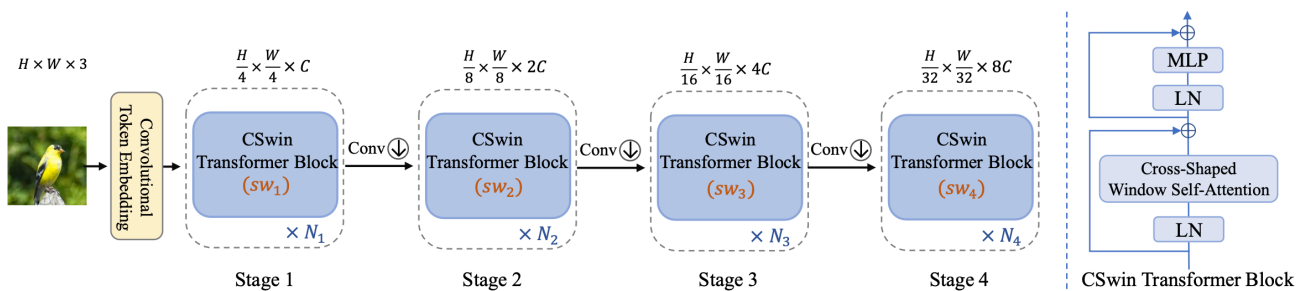
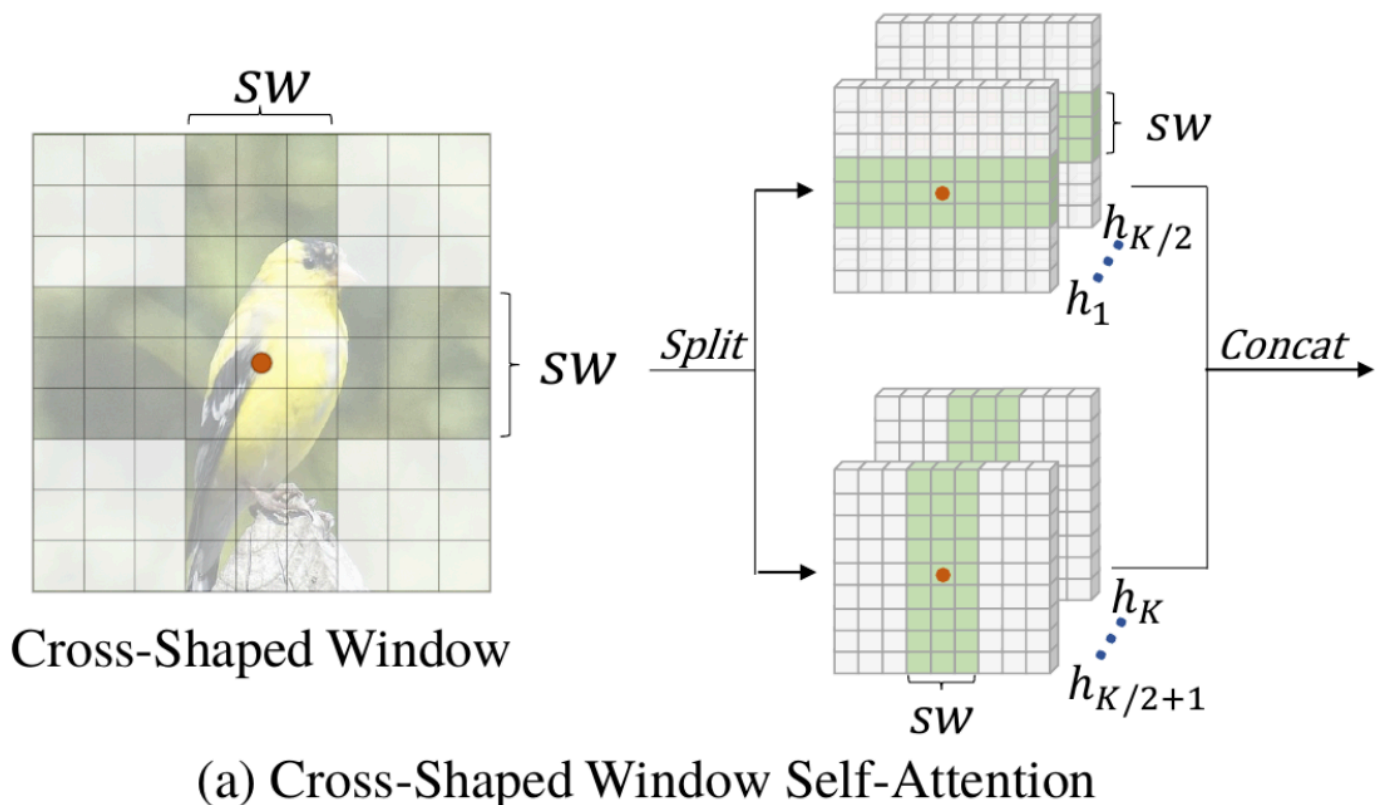


Figure 2. Left: the overall architecture of our proposed CSWin Transformer, Right: the illustration of CSWin Transformer block.

其核心是十字形状的自注意力（Cross-Shaped Window (CSWin) self-attention）。它由并行的横向自注意力和纵向的自注意力组成，对于一个多头的自注意力模型，CSWin Transformer Block 将头的一半分给和横向自注意力，另一半分给纵向自注意力，然后最后将这两个特征拼接起来，如下图所示。将设网络有 K 个头，其中 $1, \dots, K/2$ 用于横向自注意力的计算， $K/2+1, \dots, K$ 用于纵向自注意力的计算。



除此之外，作者还引入了一个能够适应不同大小输入特征的相对位置编码（Locally-enhanced Positional Encoding (LePE)），用卷积来加强局部偏置，并直接加到SA的结果后面，来减少计算量。如下图所示。

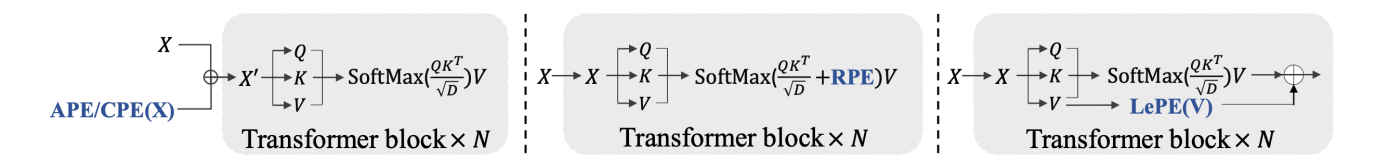


Figure 3. Comparison among different positional encoding mechanisms: APE and CPE introduce the positional information before feeding into the Transformer blocks, while RPE and our LePE operate in each Transformer block. Different from RPE that adds the positional information into the attention calculation, our LePE operates directly upon V and acts as a parallel module. * Here we only draw the self-attention part to represent the Transformer block for simplicity.

1.2 数据集

所用数据集地址：https://git.openi.org.cn/kaierlong/imagenet2012_whole/datasets

使用训练及测试数据集如下：

使用的数据集：ImageNet2012

数据集大小：共1000个类、224*224彩色图像

训练集：共1,281,167张图像

测试集：共50,000张图像

数据格式：JPEG

注：数据在dataset.py中处理。

下载数据集，目录结构如下：

└dataset

└└train

└└val

训练数据集

评估数据集

1.3 代码提交地址

暂时提交在启智中，私有未开源。

仓库地址如下：<https://git.openi.org.cn/finder4alex/CSWin-Transformer.git>

2. 代码目录结构说明

代码目录结构及说明如下：

└

└└README.md

└└README_CN.md

└└LICENSE

└└image

// 说明文档

// 中文说明文档

// 文档图片目录

```

├─ src
│   ├── args.py
│   ├── configs          // 模型参数配置目录
│   │   ├── cswin_144_24322_large_224.yaml
│   │   ├── cswin_64_24322_small_224.yaml
│   │   └─ parser.py
│   ├── data             // 数据加载及处理目录
│   │   └─ __init__.py
│   │   ├── augment
│   │   │   ├── __init__.py
│   │   │   ├── auto_augment.py
│   │   │   ├── custom_transforms.py
│   │   │   ├── mixup.py
│   │   │   └─ random_erasing.py
│   │   ├── data_utils
│   │   │   ├── __init__.py
│   │   │   └─ moxing_adapter.py
│   │   └─ imagenet.py
│   ├── models           // 模型定义目录
│   │   ├── __init__.py
│   │   ├── cswin
│   │   │   ├── __init__.py
│   │   │   ├── cswin.py      // CSWin-Transformer 网络定义文件
│   │   │   ├── get_cswin.py
│   │   │   └─ misc.py
│   ├── tools            // 相关工具目录
│   │   ├── __init__.py
│   │   ├── callback.py
│   │   ├── cell.py
│   │   ├── criterion.py
│   │   ├── get_misc.py
│   │   ├── optimizer.py
│   │   └─ schedulers.py
│   └─ trainers          // 训练目录
│       ├── __init__.py
│       ├── model_ema.py
│       ├── train_one_step_with_ema.py
│       └─ train_one_step_with_scale_and_clip_global_norm.py
├─ eval.py              // 评估文件
├─ train.py             // 训练文件
└─ train_multi.py       // 再训练 (finetune) 文件, 适配启智平台多数据集

```

3. 自验结果（交付精度规格时需要补齐）

特别说明：

- 本次训练分三阶段完成，故有三个运行环境配置。

- 超参数配置已更新为最终值，实际读者可一次完成训练，环境配置采用1即可。

3.1 自验环境

软硬件环境如下：

- 启智AI引擎： MindSpore-1.6.1-python3.7-aarch64
- Ascend: 8 * Ascend-910(32GB) | ARM: 192 核 2048GB

详细环境配置参见下图：

第一阶段环境配置：

finder4alex / CSWin-Transformer

取消关注

1

点赞

0

派生

0

代码

任务0

合并请求0

数据集

模型

云脑

项目设置

云脑 / 训练任务 / 202207271609617

2022-07-27 16:14:39 当前版本： V0001 父版本： 状态： COMPLETED 运行时长： 190:12:24

创建模型 | 修改 | 停止 | 删除

配置信息	日志	资源占用情况	结果下载
任务名称	202207271609617	AI引擎	MindSpore-1.6.1-python3.7-aarch64
状态	COMPLETED	代码分支	master
运行版本	V0001	启动文件	train.py
开始运行时间	2022-07-27 16:22:52	训练数据集	imagenet.tar.gz
运行时长	190:12:24	运行参数	arch = cswin_64_24322_small_224; run_openi = True; device_num = 8
规格	Ascend: 8 * Ascend-910(32GB) ARM: 192 核 2048GB	任务描述	
计算节点	1		

第二阶段环境配置：

finder4alex / CSWin-Transformer

取消关注1

点赞0

派生0

代码

任务0

合并请求0

数据集

模型

云脑

项目设置

云脑 / 训练任务 / 202208041495938

2022-08-04 14:57:03 当前版本: V0001 父版本: 状态: RUNNING 运行时长: 19:11:25

创建模型 修改 停止 删除

配置信息	日志	资源占用情况	结果下载
任务名称	202208041495938	AI引擎	MindSpore-1.6.1-python3.7-aarch64
状态	RUNNING	代码分支	master
运行版本	V0001	启动文件	train_multi.py
开始运行时间	2022-08-04 14:58:32	训练数据集	imagenet.tar.gz;pretrained.zip
运行时长	19:11:25	运行参数	run_openi = True; pretrained = cswin; device_num = 8; start_epoch = 319
规格	Ascend: 8 * Ascend-910(32GB) ARM: 192 核2048GB	任务描述	
计算节点	1		

第三阶段环境配置：

finder4alex / CSWin-Transformer

取消关注1

点赞0

派生0

代码

任务0

合并请求0

数据集

模型

云脑

项目设置

云脑 / 训练任务 / 202208111607247

2022-08-11 16:42:09 当前版本: V0001 父版本: 状态: RUNNING 运行时长: 17:33:49

创建模型 修改 停止 删除

配置信息	日志	资源占用情况	结果下载
任务名称	202208111607247	AI引擎	MindSpore-1.6.1-python3.7-aarch64
状态	RUNNING	代码分支	master
运行版本	V0001	启动文件	train_multi.py
开始运行时间	2022-08-11 16:42:29	训练数据集	pretrained.zip;imagenet.tar.gz
运行时长	17:33:49	运行参数	run_openi = True; pretrained = cswin; device_num = 8; start_epoch = 379
规格	Ascend: 8 * Ascend-910(32GB) ARM: 192 核2048GB	任务描述	
计算节点	1		

3.2 训练超参数

超参数配置如下：

其中data_url由启智平台实际数据地址替换，训练时替换。

特别说明：此训练超参数是支持一次训练达标的最终超参数。

```

# Architecture
arch: cswin_64_24322_small_224

# ===== Dataset ===== #
data_url: ./data/imagenet
set: ImageNet
num_classes: 1000
mix_up: 0.8
cutmix: 1.0
auto_augment: rand-m9-mstd0.5-inc1
interpolation: bicubic
re_prob: 0.25
re_mode: pixel
re_count: 1
mixup_prob: 1.0
switch_prob: 0.5
mixup_mode: batch
crop_ratio: 0.9

# ===== Learning Rate Policy ===== #
optimizer: adamw
lr_scheduler: cosine_lr
base_lr: 0.001
min_lr: 0.00001
warmup_length: 20
warmup_lr: 0.000001
cool_length: 80
cool_lr: 0.00001
tail_length: 60
tail_lr: 0.000001
over_length: 60
over_lr: 0.0000001

# ===== Network training config ===== #
amp_level: 01
keep_bn_fp32: True
beta: [ 0.9, 0.999 ]
is_dynamic_loss_scale: True
use_global_norm: True
clip_global_norm_value: 5.
enable_ema: True
ema_decay: 0.99992
loss_scale: 1024
weight_decay: 0.05
momentum: 0.9
label_smoothing: 0.1
epochs: 380

```

```
batch_size: 128

# ===== Hardware setup ===== #
num_parallel_workers: 8
device_target: Ascend

# ===== Model config ===== #
drop_path_rate: 0.4
patch_size: 4
embed_dim: 64
depth: [ 2, 4, 32, 2 ]
split_size: [ 2, 2, 7, 7 ]
num_heads: [ 2, 4, 8, 16 ]
mlp_ratio: 4.0
```

三阶段参数说明（只包含区别部分）

第一阶段参数

```
cool_length: 20
cool_lr: 0.00001
tail_length: 0
tail_lr: 0.000001
over_length: 0
over_lr: 0.0000001
epochs: 320
```

第二阶段参数

```
cool_length: 80
cool_lr: 0.00001
tail_length: 0
tail_lr: 0.000001
over_length: 0
over_lr: 0.0000001
epochs: 380
```

第三阶段参数

```
cool_length: 80
cool_lr: 0.00001
tail_length: 60
tail_lr: 0.000001
over_length: 60
over_lr: 0.0000001
epochs: 500
```

3.3 训练

3.3.1 如何启动训练脚本

训练如何启动：

- 启智平台

模型训练在启智平台完成，完整训练配置如下图所示：

参数设置:

代码分支 *

master

AI引擎

Ascend-Powered-Engine

MindSpore-1.6.1-python3.7-aarch64

启动文件 *

train.py

查看样例

数据集 *

imagenet.tar.gz

数据集位置存储在环境变量data_url中，训练输出路径存储在环境变量train_url中。

运行参数

增加运行参数

arch

cswin_64_24322_small_224

run_openi

True

device_num

8

规格 *

Ascend: 8 * Ascend-910(32GB) | ARM: 192 核 2048GB

计算节点数 *

1

新建任务

取消

- 本地命令

如果需要本地训练，可以使用如下命令：

```
python3 train.py --run_openi=True --arch=cswin_64_24322_small_224 --device_num=8
```

3.3.2 训练精度结果


- 论文精度如下：

Main Results on ImageNet

model	pretrain	resolution	acc@1	#params	FLOPs	22K model	1K model
CSWin-T	ImageNet-1K	224x224	82.8	23M	4.3G	-	model
CSWin-S	ImageNet-1k	224x224	83.6	35M	6.9G	-	model
CSWin-B	ImageNet-1k	224x224	84.2	78M	15.0G	-	model
CSWin-B	ImageNet-1k	384x384	85.5	78M	47.0G	-	model
CSWin-L	ImageNet-22k	224x224	86.5	173M	31.5G	model	model
CSWin-L	ImageNet-22k	384x384	87.5	173M	96.8G	-	model

● 复现精度如下：

○ 第一次训练结果

 [finder4alex](#) / [CSWin-Transformer](#)

取消关注

1

点赞

0

派生

0

<> 代码

① 任务

0

合并请求

0

数据集

模型

云脑

项目设置

云脑 / 训练任务 / 202207271609617

2022-07-27 16:14:39 当前版本: V0001 父版本: 状态: COMPLETED 运行时长: 190:12:24

创建模型 | 修改 | 停止 | 删除

配置信息

日志

资源占用情况

结果下载

下载日志文件

```
set_mempolicy: Operation not permitted
epoch: 0319 device: 0007 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0006 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0002 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0005 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0003 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0004 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0000 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 0319 device: 0001 acc: 0.8341746794871795, best epoch: 0319, acc is 0.8341746794871795
epoch: 320 step: 56, loss is 2.9154458045959473
epoch: 320 step: 56 loss is 2.513701084280551epoch: 320 step: 56 loss is 3.170647144317627
```

○ 第二次再训练结果

finder4alex / CSWin-Transformer

取消关注 1 点赞 0 派生 0

代码 任务 0 合并请求 0 数据集 模型 云脑 项目设置

云脑 / 训练任务 / 202208041495938

2022-08-04 14:57:03 当前版本: V0001 父版本: 状态: COMPLETED 运行时长: 37:03:48

创建模型 修改 停止 删除

配置信息 日志 资源占用情况 结果下载

下载日志文件

```
set_mempolicy: Operation not permitted
epoch: 0059 device: 0006 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0002 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0005 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0004 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0001 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0003 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0000 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 0059 device: 0007 acc: 0.835176282051282, best epoch: 0059, acc is 0.835176282051282
epoch: 60 step: 66, loss is 2.967376708984375
```

第三次再训练结果

finder4alex / CSWin-Transformer

取消关注 1 点赞 0 派生 0

代码 任务 0 合并请求 0 数据集 模型 云脑 项目设置

云脑 / 训练任务 / 202208111607247

2022-08-11 16:42:09 当前版本: V0001 父版本: 状态: RUNNING 运行时长: 17:30:45

创建模型 修改 停止 删除

配置信息 日志 资源占用情况 结果下载

下载日志文件

```
set_mempolicy: Operation not permitted
epoch: 0027 device: 0005 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0002 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0006 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0000 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0001 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0003 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0004 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 0027 device: 0007 acc: 0.8358173076923077, best epoch: 0026, acc is 0.8360576923076923
epoch: 28 step: 98, loss is 2.5100274085998535epoch: 28 step: 98, loss is 2.7879655361175537epoch: 28 step: 98, loss is 3.028127670288086
```

精度结果对比

- 论文精度为: 83.6
- 复现精度为:
 - 83.42 (第一阶段最优值)
 - 83.51 (第二阶段再训练最优值)
 - 83.60 (第三阶段再训练最优值)
- 对比结论: 达到论文精度

3.4 模型推理

推理命令如下：

```
python3 eval.py --config=src/configs/cswin_64_24322_small_224.yaml --pretrained={ckpt_path} --device_id={device_id} --device_target={device_target} --data_url={data_url}
```

4. 参考资料

4.1 参考论文

- [CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows](#)

4.2 参考git项目

- [microsoft/CSWin-Transformer](#)

4.3 参考文献

- [CSWin-T: 微软、中科院提出十字形注意力的 CSWin Transformer](#)
- [Vision Transformer 之 CSWin Transformer](#)