# 210901 训练又不收敛排查@hp

背景：洪鹏21年8月份基于<https://30.16.73.65:30000/seg/mmsegmentation/commits/lalei>，commits==04fc9b73c1597b792259aa241b421fd5fe226394("[add repvgg repseg ddrnet, fix some test problem](https://30.16.73.65:30000/seg/mmsegmentation/commit/04fc9b73c1597b792259aa241b421fd5fe226394)"), 以及commits==80590e11fdcc0ba0cc699768c13fb6cbffbc2563("[final bmseg->mmseg code, erf1head miou on mmseg is greater than bmseg, erf2head is almost same!](https://30.16.73.65:30000/seg/mmsegmentation/commit/80590e11fdcc0ba0cc699768c13fb6cbffbc2563)")训练多次

<https://30.16.73.65:30000/seg/mmsegmentation/blob/lalei/configs/erfnet/erfnet_v0_catbn_iter_600e_guass_mmseg75.47_nobndropout_900e_sgd0.01-to-sgd0.02_resize0.8-1.25_pool-to-conv1x1.py>均发散。

我训练了多次也发散。

# baseline:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据集：  train\_data: train\_dataset\_list.txt, 6839  train\_data: val\_dataset\_list.txt, 760  环境：  Python: 3.7.10  Torch: 1.6.0  TorchVision: 0.7.0  OpenCV: 4.5.1  MMCV: 1.2.7  MMCV Compiler: GCC 8.3  MMCV CUDA Compiler: 10.2  MMSegmentation: 0.11.0+  基本配置：batch\_size=16, gpu=2, worker\_per\_gpu=4, sampler\_per\_gpu=8 | | | | | |
| 框架 | mIoU | 收敛性 | 编号 | 对照试验 | 试验参数 |
| erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47.py | | | | | |
| mmseg | 0.7547 | 稳定收敛 | EXP11 | EXP6 | backbone：ERFNet\_Gov(bn在cat之后)，  head: ERFHead\_Gov+Aux dropout\_ratio=[0.1, 0],aux\_align=1  数据AUG：GaussBlur+crop{fg/(fg+bg)>0.5}+Resize(0.5,2.0)  weight=(5e-4) —>(2e-4)  其它同"EXP6" |
| erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_nobndropout\_900e\_sgd0.01-to-sgd0.02\_resize0.8-1.25\_pool-to-conv1x1.py | | | | | |
| mmseg | 0.7732 | 稳定收敛 | EXP2  EXP16+  EXP22+  EXP30\_3+2head'  EXP9' | EXP11 | 考虑收敛时间及收敛精度，做如下组合：  1）resize(0.5,2)—>resize(0.80,1.25)  2）600e->900e  3) downsampler\_bottleneck无pool，全部用1\*1/s2的conv代替pool, conv后紧跟bn  4）sgd: 0.01->0.02，  5) 3epoch warmup  其它同"EXP11" |

# 复现过程记录：

基本环境除了MMCV Compiler: GCC稍微有点差异以外，其它均保持一致。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| baseline:  数据集：  train\_data: train\_dataset\_list.txt, 6839  train\_data: val\_dataset\_list.txt, 760  环境：  Python: 3.7.10  Torch: 1.6.0  TorchVision: 0.7.0  OpenCV: 4.5.1  MMCV: 1.2.7  MMCV Compiler: GCC 8.4  MMCV CUDA Compiler: 10.2  MMSegmentation: 0.11.0+  基本配置：batch\_size=16, gpu=2, worker\_per\_gpu=4, sampler\_per\_gpu=8 | | | | | |
| 框架 | mIoU | 收敛性 | 编号 | 对照试验 | 试验参数 |
| erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1.py | | | | | |
| mmseg.8059 | 75.25 | 收敛 | EXP11.0 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1.py  1、backbone=ERFNet\_Gov\_v0  2、有dropout  3、Resize=(0.5, 2.0)  4、lr=0.01  5、max\_iters=256200  6、无warmup |
| mmseg.8059 | 75.66 | 收敛 | EXP11.1 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_dp\_lr0.01\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_v0  2、有dropout  3、Resize=(0.5, 2.0)  4、lr=0.01  5、max\_iters=384300  6、有warmup  分析：  1、迭代次数和warmup不影响收敛性 |
| mmseg.8059 | 77.04 | 收敛 | EXP11.2 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_dp\_lr0.02\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_v0  2、有dropout  3、Resize=(0.5, 2.0)  4、lr=0.02  5、max\_iters=384300  6、有warmup  分析：  1、学习率不影响收敛性 |
| mmseg.8059 | 77.23 | 收敛 | EXP11.3 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_dp\_lr0.02\_warmup\_3843\_resize0.8-1.25.py  1、backbone=ERFNet\_Gov\_v0  2、有dropout  3、Resize=(0.8, 1.25)  4、lr=0.02  5、max\_iters=384300  6、有warmup  分析：  1、Resize不影响收敛性 |
| mmseg.8059 |  | 发散 | EXP11.4 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_bb\_dp\_lr0.02\_warmup\_3843\_resize0.8-1.25.py  1、backbone=ERFNet\_Gov\_Pool2Conv1x1  2、有dropout  3、Resize=(0.8, 1.25)  4、lr=0.02  5、max\_iters=384300  6、有warmup  分析：  1、backbone与该组合的Resize会导致发散 |
| mmseg.8059 | 76.78 | 收敛 | EXP11.5 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_bb\_dp\_lr0.02\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_Pool2Conv1x1  2、有dropout  3、Resize=(0.5, 2.0)  4、lr=0.02  5、max\_iters=384300  6、有warmup  分析：  1、backbone与该组合的Resize收敛 |
| mmseg.8059 |  | 发散 | EXP11.6 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_nodp\_lr0.01\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_v0  2、无dropout  3、Resize=(0.5, 2.0)  4、lr=0.01  5、max\_iters=384300  6、有warmup |
| mmseg.8059 |  | 发散 | EXP11.7 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_nodp\_lr0.02\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_v0  2、无dropout  3、Resize=(0.5, 2.0)  4、lr=0.02  5、max\_iters=384300  6、有warmup |
| mmseg.8059 |  | 发散 | EXP11.8 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_bb\_nodp\_lr0.02\_warmup\_3843.py  1、backbone=ERFNet\_Gov\_Pool2Conv1x1  2、无dropout  3、Resize=(0.5, 2.0)  4、lr=0.02  5、max\_iters=384300  6、有warmup |
| mmseg.8059 |  |  | EXP11.9 | EXP11 | erfnet\_v0\_catbn\_iter\_600e\_guass\_mmseg75.47\_fuxian1\_modify\_nodp\_lr0.02\_warmup\_3843\_resize0.8-1.25.py  1、backbone=ERFNet\_Gov\_v0  2、无dropout  3、Resize=(0.8, 1.25)  4、lr=0.02  5、max\_iters=384300  6、有warmup |
| 1、确认是否没有dropout就是发散  2、目前看来EXP11.5是最佳的实验配置 | | | | | |

1. 标注数据的banma\_annotation未经过核验，直接入库，错误大概率来源于这里。可以去掉banma\_annotation跑一个模型看看是否收敛

1. topK Error, K=50, 错误的图片直接删掉，训练模型看是否收敛。

1. 排除由于label错误导致的loss异常

loss: 0.2 70+

loss: 0.3 60+

loss: 0.8 20+

0.3突然升到loss: 0.8199 or loss: 0.850

random操作，resize\_aug+crop\_aug+batch\_size

* 1. 简单可视化观察采样的样本。模型没关系。

* 1. **loss异常的单张采样样本找出来：**

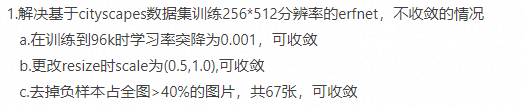
* + 1. **name, scale, crop区域的loss。推荐。已复现**

* + 1. **图片+loss。不推荐**

1. 排查除了label以外的导致模型发散的原因。

* 1. erfnet+cityscape，训练分辨率512\*256, 发散了。找小莉咨询一下经验。

* + 1. 看下扣的样本，加条件fg/(fg+bg)后也收敛。



1. 样本ADS清洗。

1. 可以参阅一下learning from noise label的paper的做法，提高噪声样本的学习稳定性。

[Learning with Noisy Labels](http://papers.neurips.cc/paper/5073-learning-with-noisy-labels.pdf)

[Learning from Noisy Labels with Deep Neural Networks](https://arxiv.org/abs/2007.08199)

A Topological Filter for Learning with Label Noise

Transfer Learning with Label Noise