

DL HW2

任伟宁 汽 75 2017010672

April 21, 2021

TaskA

(1)Run the model

训练曲线如图1所示

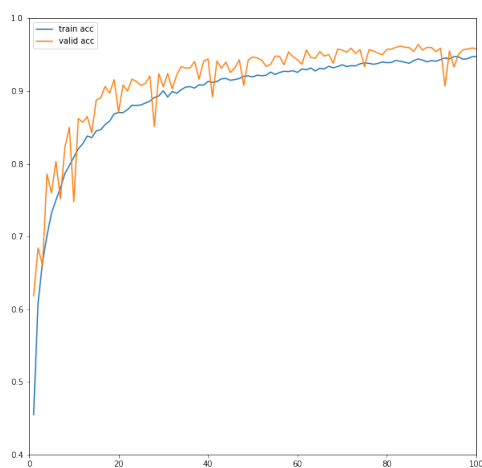


图 1: Task A 训练曲线

在 validation set 上准确率为 0.9630。

(2) 可视化

使用 t-sne 可视化结果如图2所示

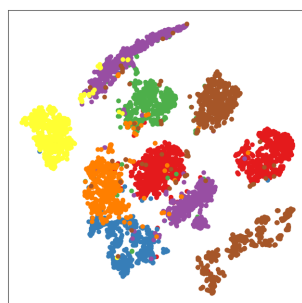


图 2: tsne 可视化效果

我选择的是 resnet50 layer4 中最后一个 1x1 卷积核进行的可视化，分别选择了 Pasture 和 Industrail 进行可视化。两种方法是将输入图片前向传播到相应层，之后在相应层上最小化均值（最大化特征）后反向传播，于是可以在原图片上观察到相应层提取到的特征。首先分析 Pasture_1.png 上对比两种方法结果。

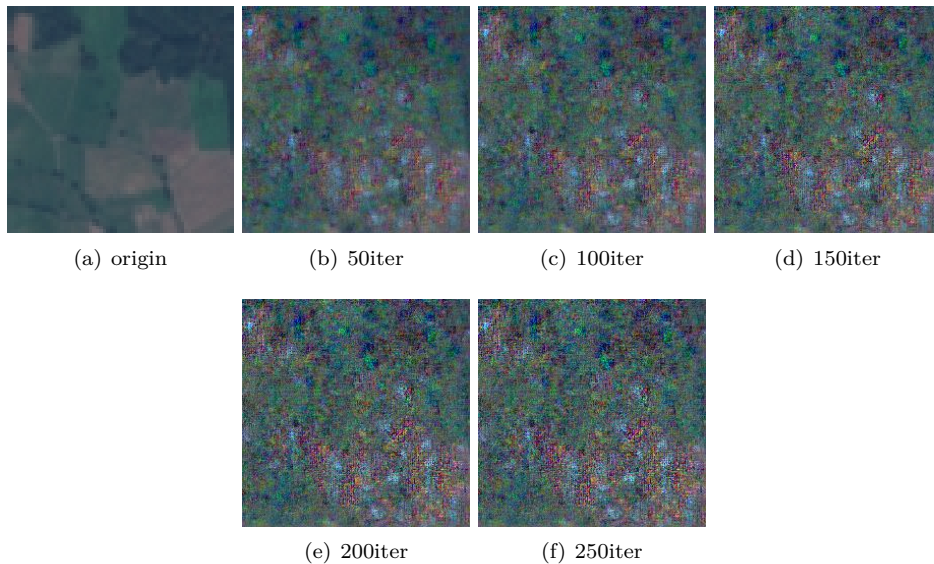


图 3: deep dream 下的 pasture_1.jpg

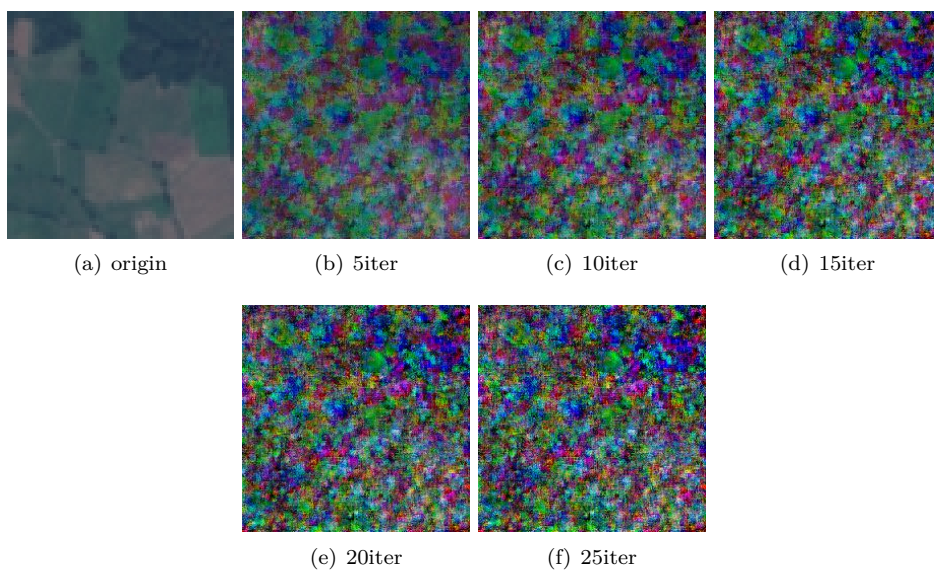


图 4: cnn visualization 下的 pasture_1.jpg

接下来分析 Industrial_01.jpg 可视化效果

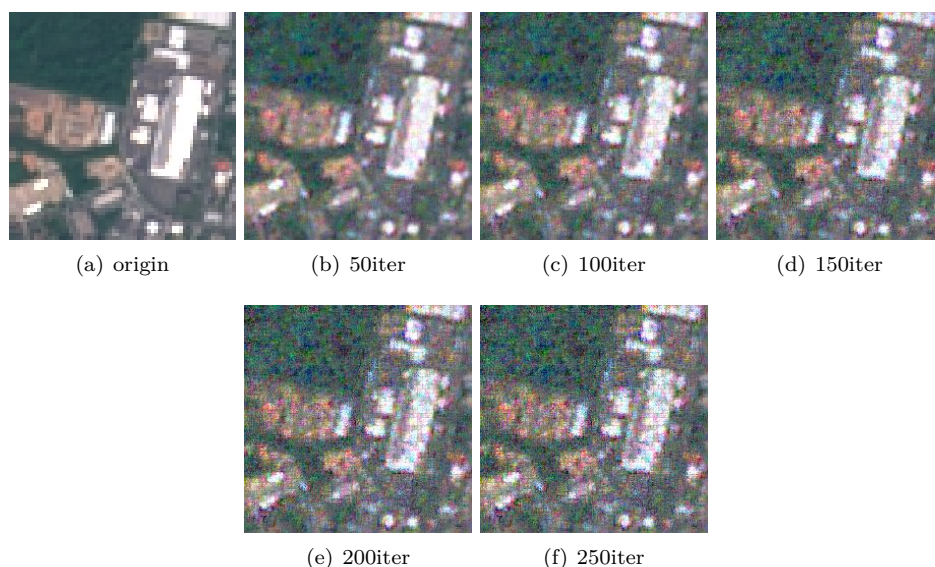


图 5: deep dream 下的 Industrial_1.jpg

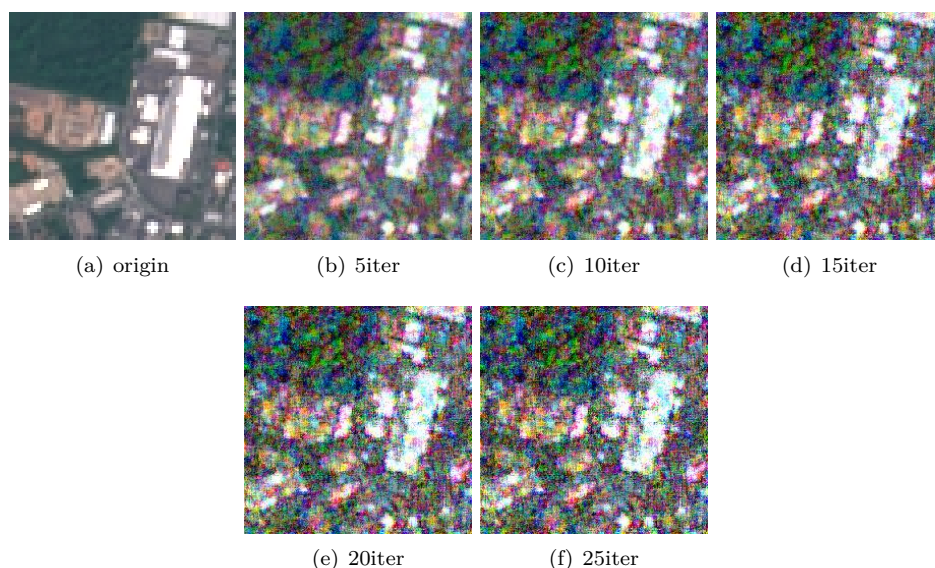


图 6: cnn visualization 下的 Industrial_1.jpg

从网上的一些例子上可以看出，deep dream 更倾向与捕捉边缘特征或者局部的特征（如眼睛），之后在原有图片上如同加水印一般反复出现，从我的图像中可以看出有 Industrail 的图像中有周期性出现方块的趋势，说我猜测或许捕捉到了一些小白顶工厂的局部特征，并对其反复出现。，但是对颜色的改变并不明显。从我复现的例子上也看出了一些这个端倪但是可能因为数据集上局部特征有限，且分辨率相对较低，没有复现出网上的效果。而对 cnn_visualization 来说，它仿佛更倾向于提取颜色特征，得到了很鲜明的颜色特征。仔细观察图 6 中草地部分，如果原本有一定纹理，可视化后出现紫色的可能性更大。

Task B

模型设计

我首先比较了 resnet18,resnet34,resnet50 三个网络的表现作为 baseline，发现 resnet50 的效果明显不如其余两个网络。在未调整学习率，采用 SGD 优化器下三者 valid 曲线如图所示。

由于数据集大小有限，更深的 resnet50 或许并不适合这个工作，考虑到 34 和 18 精度几乎一

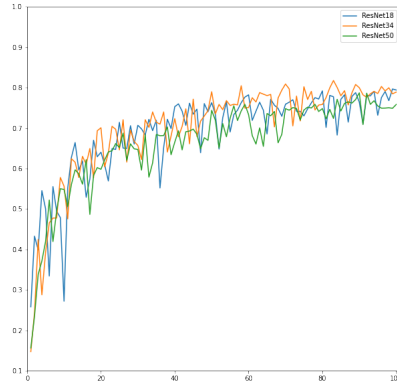


图 7: B Baseline

样，valid set 上大概都是 0.80，我于是选择了 18 作为我设计自己模型的基础。我想到老师上课讲到，网络的深度和宽度都很重要，于是我选择对 resnet 的宽度进行增加，设计了有两条通路的 resnet，我称之为 DLResNet (Double Layer)，考虑到若采用一样的 3x3 核只等价与增加了 channel 数量，我在另外一个通道选择了不同的卷积核。我对 BasicBlock 的修改如下图所示。

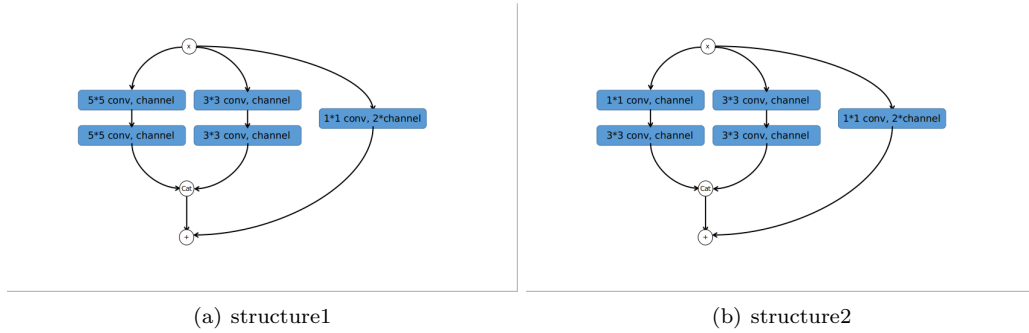


图 8: 设计的两种结构

在后续实验中，发现两种结构基本没有区别，都可以大概提高两个百分点的精度，达到 0.81 ~ 0.82，将 SGD 改为 Adam 后又可以提升一个点的精度。说明双层结构有效，这启示我或许增加通道数可以提高模型的性能，于是我思考若采用类似 resnext 模型，增加更多通道数，是否可以获得更高的精度呢？事实证明这种方案是有效的，resnext50 在 100 个 epoch 后 acc 达到了 0.86。于是我对 resnext50 和我的双层结构均进行了学习率的调整，训练了 200 个 epoch，二者训练曲线 (valid acc) 如图9。可以看到 ResNext50 显著好于我的 DLResNet

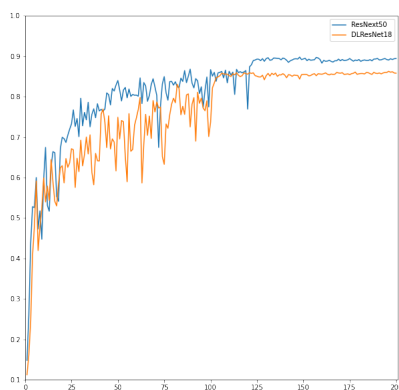


图 9: ResNext50 vs DLResNet18

考虑到在 resnet 在 baseline 测试中 50 相比 18 相差较大, 我想测试一下 resnext18 的表现如何。torch 实现的 Resnet 版本中并没有 resnext18 的实现, 于是我在原本基础上加入了 group conv。发现表现和 resnext50 基本没有区别, 精度上只差了 0.2~0.3 个点。二者训练曲线对比如图10。最后我的 ResNext18 在 valid set 上的 acc 是 0.8906。

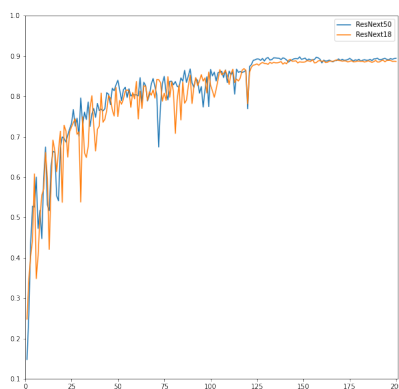


图 10: ResNext50 vs ResNext18

在学习率策略的调整上, 我最后采用 MultiStepLR 的方法, milestone 设为 [80,120,160], 初始学习率 0.001, 每次变为 0.2 的策略。我也尝试过 cos 和 exp 调整的方式, 感觉最后大概都是 0.88~0.89 之间, 没有本质改变。但是每一次学习率跳变的过程, 可以看到 acc 会提高几个百分点。

接下来对数据增强和学习率策略进行 ablation study。None strategy, only Data Argument, only LR adjust 以及 both 的 valid acc 曲线如下图11所示。

可以看到数据增强以及学习率策略都是有效的。加入每一个策略大概都可以提高两个点的精度。合在一起也可以提升精度。

由于原本代码已经提供了 RandomCrop 和 RandomFlip 两个数据增强的功能, 我在做无数据增强的实验时去掉了二者。在做有数据增强的实验中我除了二者加入了 ColorJitter。

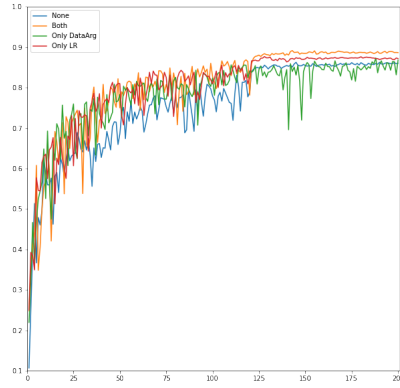


图 11: Ablation study

TaskC: Long-Tailed

由于不清楚引入预训练好的 ResNet18 是否属于引入数据，这部分我分别使用了预训练好的 ResNet18 和自己的 ResNext18 进行了测试。预训练好的 ResNet18 acc 可以达到 0.8959。我自己在 TaskB 中搭建的模型从头开始训可以达到 0.8006，其训练曲线如图12。可以看到由于尾分布在

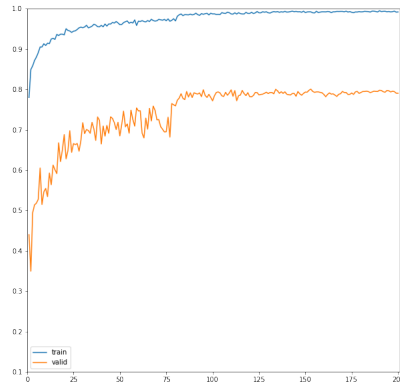


图 12: Task C1

Train Set 上和 Valid Set 上的分布不同，所以两条曲线区别明显。图13为 confusion matrix。从中可以看出，网络有把尾分布分类为头分布的倾向，其中 4、6、7 类尤为明显，他们在训练集中只出现了 19,10,36 次。

```
validation: 1.0650, 0.8006
Confusion Matrix is :
[[566 3 0 0 1 1 0 0 1 1]
 [ 0 570 0 0 0 0 0 0 0 0]
 [ 11 12 559 1 1 0 4 3 0 0]
 [ 54 5 0 267 47 2 13 54 72 0]
 [ 0 0 2 2 470 0 6 34 2 0]
 [128 60 18 10 0 109 34 6 64 0]
 [110 2 85 0 12 1 263 40 8 0]
 [ 0 0 1 0 3 0 0 573 0 0]
 [ 51 7 16 13 3 1 3 4 489 0]
 [ 37 7 2 0 0 0 0 0 11 537]]
```

图 13: Confusion Matrix

(2)

我采用了 2020NIPS 上 [1] 的 balanced softmax 进行重加权。文中同样的另一贡献是引入了 meta-sampler, 不过作者说它是为了对待 extreme imbalance 场景设计的，比如某一类只出现一次，

故我这里没有使用（另外一个原因是看不懂论文）。

作者两个角度分析引入 balanced softmax 的 motivation。其一是对于长尾分布来说，train set 和 valid set 之间存在 label distribution shift。即对于无偏的 valid set 上，从贝叶斯推断的角度来讲，有

$$\phi_j = p(y = j | x) = \frac{p(x | y = j)p(y = j)}{p(x)} \quad (1)$$

而当我们到了存在 label shift 的 train set 上时，由于 $p(y = j) p(x)$ 两项估计不准确。于是作者提出这样一个 Theorem，即假设在不平衡的数据集上我们要计算的概率为 $\hat{\phi}_j$ ，其贝叶斯估计为 $\hat{\phi}_j = \hat{p}(y = j | x) = \frac{p(x|y=j)}{\hat{p}(x)} \frac{n_j}{\sum_{i=1}^k n_i}$ 。如果 ϕ_j 用 softmax 度量的话，那么简单进行等式变换即有

$$\hat{\phi}_j = \frac{n_j e^{\eta_j}}{\sum_{i=1}^k n_i e^{\eta_i}} \quad (2)$$

同时，作者还从泛化误差界的角度分析了 balanced softmax(BALM) 可以最小化测试集上的泛化误差界，其中涉及了一些证明，这里不再展开。

在加入 BALM 之后，我在自己的 ResNext18 和预训练的 ResNet18 上分别进行了测试，在 valid set 上 acc 分别提升了约 4.5 个百分点和 8 个百分点，达到了 0.9404 和 0.8796 其测试 acc 曲线如图14所示。在 resnet18 上我只训练了 100 个 epoch，因为感觉效果已经很好了。

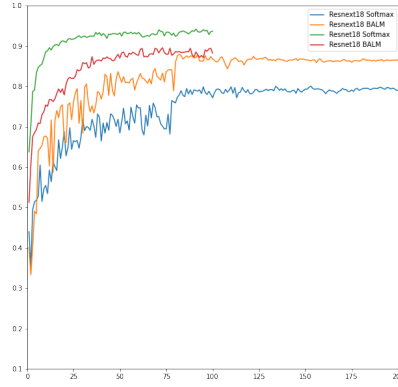


图 14: Task C2

References

- [1] J. Ren, C. Yu, S. Sheng, X. Ma, H. Zhao, S. Yi, and H. Li, “Balanced meta-softmax for long-tailed visual recognition,” in *Proceedings of Neural Information Processing Systems(NeurIPS)*, Dec 2020.