论文笔记：Fixing the train-test resolution discrepancy

原创 Zerg_Wang 于 2020-06-07 16:25:59 发布 1690 收藏 3 原力计划 编辑 版权

分类专栏： Machine Learning 文章标签： 深度学习 机器学习 数据增强 image augment 神经网络

前言

本文仅个人理（nao）解（bu），如有错误，欢迎指正！

论文地址：https://arxiv.org/pdf/1906.06423v3.pdf，该文章被NeurIPS 2019收录。

该文章探究了在深度学习的图像分类任务中，训练及测试图像分辨率对模型的影响。作者给出了自己的结论：a lower train resolution improves the classification at test time!

简介

为了使模型获得较好的性能，训练数据和测试数据在分布上应该是一致的。然而，针对训练数据和测试数据的一些操作，如数据集的制作、训练前的预处理、数据增强等，是不一致的。例如，针对训练集中的图像，研究者可能会先从中随机裁剪一块矩形图像区域，并通过进一步的裁剪、放缩使其符合模型的输入要求。而对于测试集中的图像，往往只是简单的中心裁剪（CenterCrop）。作者认为这会导致训练数据和测试数据在缩放尺度上的不同，从而导致两者数据分布不一，最终影响模型性能。

对于以上的问题，以往研究提出了一些基于更改分辨率的解决思路。例如：

- 数据增强，通过将不同分辨率（或者说，不同缩放尺寸）的图像输入网络，使网络学习学习到这种尺度的不变性。
- feature pyramid net，通过网络将图像缩放成不同分辨率进行学习。

不同于以上工作，作者提出自己的解决方法：应该对训练和测试数据加以联合优化，缩小训练集的分辨率，或者增大测试集的分辨率，从而使它们的数据分布相符合。而对于该联合优化所带来的分布上的变化，辅以网络结构上的微调，可以带来更好的模型性能。

增强及预处理的影响

此处探讨前文所述的区别对待训练集和测试集的处理策略，以及具体预处理或增强方式（RandomResizedCrop、CenterCrop）的影响。

RoC（Region of Classification）

RoC是输入图像的一部分，一般为一块矩形图像，一般在数据增强、预处理等过程中从输入图像中取得。注意：RoC图像不一定

然而，前文提到，针对训练集和测试集，有不同的数据增强或预处理策略，因此训练集和测试集的RoC基本不同。例如，对于测试集图像使用中心裁剪，对训练集使用随机裁剪缩放（RandomResizedCrop）后，这两部分的RoC所占原输入图像比例的概率分布情况：

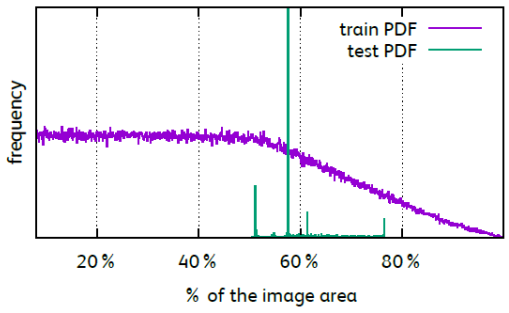

Figure 2: Empirical distribution of the areas of the RoCs as a fraction of the image areas extracted by data augmentation. The data augmentation schemes are the standard ones used at training and testing time for CNN classifiers. The spiky distribution at test time is due to the fact that RoCs are center crops and the only remaining variability is due to the different image aspect ratios. Notice that the distribution is very different at training and testing time.

可见训练集数据与测试集数据分布完全不同。

目标大小

针对训练集的预处理或增强方式：对于原图进行Pytorch自带RandomResizedCrop后，再通过指定的缩放因子s缩放到指定大小Ktrain × Ktrain。

针对测试集的预处理或增强方式：将原图按缩放因子s缩放后再CenterCrop，得到指定大小Ktest × Ktest

测试集中的目标大小，为训练集中的80%。

对网络结构的影响

训练集输入图像分辨率固定为224×224，不同分辨率的测试集图像输入网络后，网络中激活层和池化层的结构也将不同，网络输出结果也会不同，下图展示了不同分辨率的图像输入下，average pooling层输出结果的分布函数：

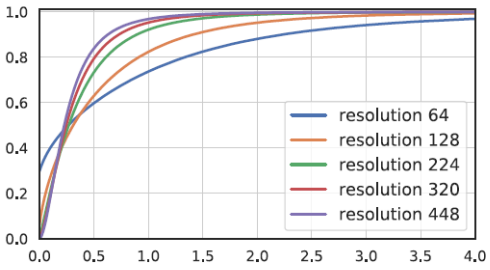

Figure 3: Cumulative density function of the vectors components on output of the spatial average pooling operator, for a standard ResNet-50 trained at resolution 224, and tested at different resolutions. The distribution is measured on the validation images of Imagenet.

当测试集图像分辨率从224降为64时，ReLU激活层尺寸（activation map）从7×7变为2×2，这意味着其输出值为0的可能性变高了，输出值在分布上也变得更为稀疏，从而直接影响到池化层的输出。相反，提升测试集图像的分辨率，则输出的分布会相对集中。

对精度的影响

K_{test}	64	128	224	256	288	320	384	448
accuracy	29.4	65.4	77.0	78.0	78.4	78.3	77.7	76.6

同样是保持Ktrain=224，探究不同Ktest对精度的影响，可见当Ktest=288时精度最高。前文作者证明了在默认的预处理或增强下，测试集中的目标大小为训练集的80%，因此将Ktest扩大到原来的1.25倍后（224×1.25=280），训练集和测试集的目标大小将大致相同，这将减少CNN对尺寸不变性的学习（作者前文提到过，CNN对于尺寸的变化是不敏感的，换言之，CNN难以学习到尺寸不变性），从而实现性能的提升。

改进方法

正如前文所述，对输入图像分辨率的更改的确起到作用，但由于其扭曲了数据的分布，网络可能无法通过数据的“本来面貌”去学习，对此，作者提出以下补偿策略：

- Parametric adaptation：测试时，在网络的pooling层后，作者提出加入一个equalization操作，通过标量变换（scalar transformation）将数据分布调整为原分辨率大小的数据分布。但作者认为该做法效果有限，因为网络过于简单，难以区分经过pooling后的不同的数据分布。
- Fine-tuning：通过对测试数据分布的分析，作者发现，应该对分布的稀疏性进行调整，这需要在pooling前引入批归一化操作并进行微调。通过下图可以看出，微调后，数据在经过average pooling后的分布情况与训练数据的分布更为接近。然而，作者也提到，这些做法只是对于数据分布的一个修正，不一定能带来精度的提升。

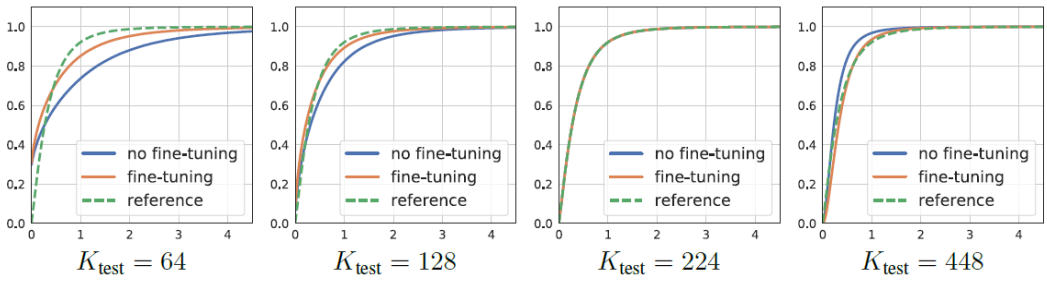

Figure 4: CDF of the activations on output of the average pooling layer, for a ResNet-50, when tested at different resolutions K_{test}. Compare the state before and after fine-tuning the batch-norm.

实验结果

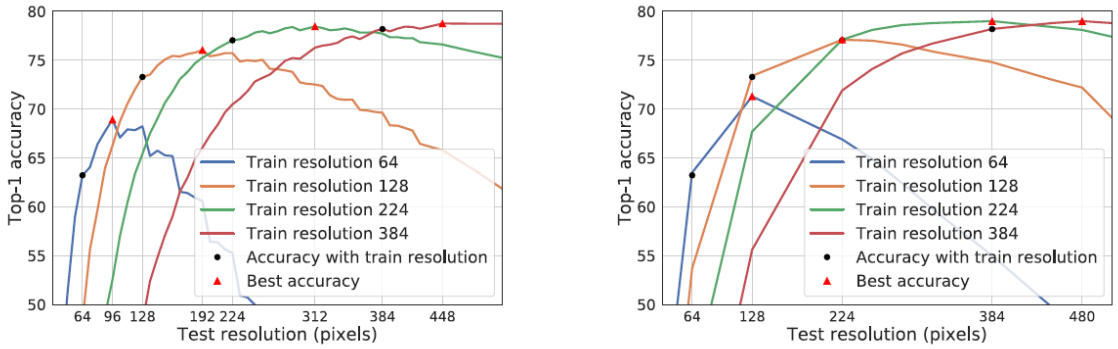

Figure 5: Top-1 accuracy of the ResNet-50 according to the test time resolution. Left: without adaptation, right: after resolution adaptation. The numerical results are reported in Appendix C. A comparison of results without random resized crop is reported in Appendix D.

右图展示的是对最后一层的批归一化层微调后的结果，可见fine-tune对性能有一定的正向影响。同样，fine-tune令达到最佳性能所需要的测试集分辨率更高了，当训练集分辨率为224时，使用fine-tune使达到最佳精度的测试集分辨率从288增加到384。

Table 1: Application to larger networks: Resulting top-1 accuracy.

Model	Train	Fine-tuning			Test resolution					
	resolution	Class.	BN	3 cells	331	384	395	416	448	480
PNASNet-5-Large	331	_	_	_	82.7	83.0	**83.2**	83.0	83.0	82.8
PNASNet-5-Large	331	✓	✓	_	82.7	83.4	83.5	83.4	**83.5**	83.4
PNASNet-5-Large	331	✓	✓	✓	82.7	83.3	83.4	83.5	83.6	**83.7**
		Class.	BN	3 conv.	224		288		320	
ResNeXt-101 32x48d	224	✓	✓	_	85.4		86.1		**86.4**	

Table 3: Transfer learning task with our method and comparison with the state of the art. We only compare ImageNet-based transfer learning results with a single center crop for the evaluation (if available, otherwise we report the best published result) without any change in architecture compared to the one used on ImageNet. We report the Top-1 accuracy (%).

Dataset	Models	Baseline	+our method	State-of-the-art models	
Stanford Cars [16]	SENet-154	94.0	94.4	EfficientNet-B7 [34]	**94.7**
CUB-200-2011 [36]	SENet-154	88.4	**88.7**	MPN-COV [19]	**88.7**
Oxford 102 Flowers [23]	InceptionResNet-V2	95.0	95.7	EfficientNet-B7 [34]	**98.8**
Oxford-IIIT Pets [25]	SENet-154	94.6	94.8	AmoebaNet-B (6,512) [14]	**95.9**
Birdsnap [4]	SENet-154	83.4	**84.3**	EfficientNet-B7 [34]	**84.3**

至于该实验在ImageNet上达到的各种SOTA性能的阐释，详见论文，这里就不赘述了。