深度学习工作站装机指南

2018-12-16作者[蒂姆·德特默斯](https://timdettmers.com/author/tim-dettmers/)

深度学习的计算量非常大，需要配多个内核的快速CPU吗？买快的CPU会不会太浪费？搭建深度学习系统时，在不必要的硬件上浪费钱就太糟了。在这里，我将引导您逐步了解物美价廉的系统所需的硬件。

这些年来，我总共装了7个不同的深度学习工作站，尽管进行了认真的研究和分析，但我在选择硬件配件的时候还是犯了不少错误。在本指南中，我想分享我多年来的经验，以便您不会犯我以前犯过的错误。

该博客文章按错误严重性排序。让我们从最费钱的错误开始。

# 显卡

本文假定您将使用GPU进行深度学习。如果您要搭建或升级用于深度学习的系统，GPU是个绕不开的话题。GPU是深度学习应用的心脏, 算力的作用实在是太大了，不容忽视。

我在博文[我的GPU建议](http://timdettmers.com/2018/11/05/which-gpu-for-deep-learning/)中详细讨论了GPU的选择，选择GPU可能是深度学习系统最关键的选择。选择GPU时，最容易犯的错误有三个：（1）性价比不高；（2）显存不足；（3）散热不好。

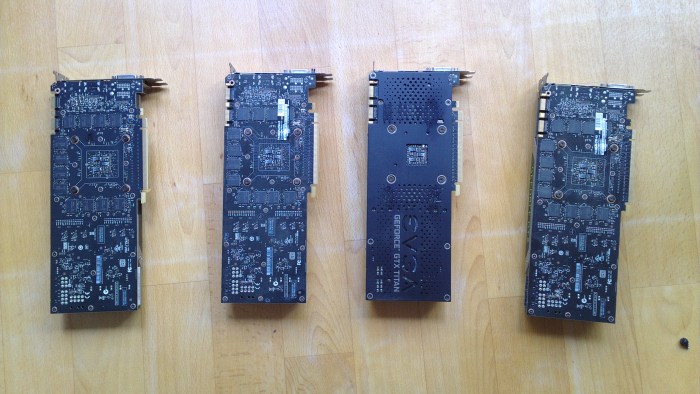
为了获得好的性价比，我建议使用RTX 2070或RTX 2080 Ti。如果使用这些卡，则应使用16位(数据宽度)的模型。另外， GTX 1070，GTX 1080，GTX 1070 Ti和GTX 1080 Ti也是不错的选择，只是在这些卡上, 不能使用16位模型, 只能使用32位模型。

选择GPU时，请注意显存要求。因为RTX卡能跑16位模型, 而16位模型需要的内存少一半, 所以在同样的显存配置下, RTX能跑的模型大小是GTX卡的两倍。考虑到RTX卡的这项优势，选用RTX卡, 并学习如何有效使用16位模型, 是非常有意义的。通常，对显存的需求大致如下：

* 追求最高水平分数的研究：>=11 GB
* 探寻有趣架构的研究：>=8 GB
* 其他的研究：8 GB
* Kaggle：4 – 8 GB
* 初学者：8 GB (但是要注意确认下应用领域的模型大小)
* 企业：8 GB 用于原型， >=11 GB 用于训练

另一个需要特别注意的问题是散热。如果插GPU的PCIe插槽挨得太近，则应该使用鼓风机式风扇的GPU (如下图, 热气被显卡直接排出到机箱的后面) 。否则可能遇到过热问题，这会导致GPU速度变慢（大约30％）并过早老化。



[](https://i2.wp.com/timdettmers.com/wp-content/uploads/2015/03/suspectlineup.jpg)

**是什么出了问题?**  
当性能不佳时, 您能识别出出现故障的配件吗？是某个GPU？还是CPU的问题？

# 内存

内存方面容易犯的错误, 主要是内存主频太高。另外就是内存太小，导致原型构建不够流畅。

## 所需的内存主频

内存主频只是营销的噱头，内存公司误导用户购买“更快的” 内存，但实际上根本没什么用。这在 Linus Tech Tips的视频" [内存速度真的重要吗？](https://www.youtube.com/watch?v=D_Yt4vSZKVk)”里面解释的很清楚。

另外很重要的一点是, 内存的速度和内存与显存之间的传输几乎无关。这是因为（1）如果使用页锁定内存(pinned memory)，那么小批量(mini-batch)处理将被转移到GPU，而不会涉及CPU；（2）如果不使用页锁定内存，则快速内存比慢速内存大概只快0-3％, 所以, 把钱花在其他地方！

## 内存大小

内存大小不会影响深度学习性能。但是，这可能会拖慢GPU代码执行（内存足够时, 无需交换数据到磁盘）。内存要足, 不能影响GPU。内存量应该至少和显存最大的GPU相当。例如，如果您的Titan RTX具有24 GB的内存，则应至少配上24 GB的内存。不用因为有多张GPU而加多内存。

这种“内存匹配最大显存”策略的问题在于，如果要处理大型数据集，可能仍然缺少内存。最好的策略是匹配您的GPU，当然, 如果您觉得内存不足，那就多买点。

另一种策略和心理学有关：心理学告诉我们，注意力是一种随着时间而消耗的资源。内存是少数几个可以节省更多精力来解决更困难编程问题的配件之一。如果有更多的内存，可以将精力集中在更紧迫的事情上，而不是花很多时间来解决内存瓶颈。有了大量的内存，可以避免这些瓶颈，节省时间并在遇到更紧迫的问题时提高效率。特别是在Kaggle比赛中，我发现额外的内存对于功能设计非常有用。因此，如果您有足够的钱并需要进行大量预处理，那么额外的内存可能是一个不错的选择。因此，按照这个思路，先把内存买回来, 免得到用时方恨少。

# 中央处理器(CPU)

这方面容易犯的主要错误是大家过分关注CPU的PCIe通道了。不必太在意PCIe通道。相反，只要确认CPU和主板组合是否支持您要运行的GPU数量即可。第二个最常见的错误是买的CPU太强大了。

## CPU和PCI-Express

大家为PCIe通道疯狂！但是，事实上它对深度学习性能几乎没什么影响。如果您只有一个GPU，则仅需要PCIe通道即可将数据从CPU 内存快速传输到显存。但是，包含32个32位图像（32x225x225x3）的ImageNet批量, 在16通道时需要1.1毫秒，8通道时2.3毫秒,4通道4.5毫秒。这是理论数字，在实践中，PCIe速度通常会慢一倍, 但这仍然是闪电般的快！PCIe通道通常具有纳秒级的延迟，因此可以忽略延迟。

综合起来，对于32个图像的ImageNet小批量和一个ResNet-152，其用时如下：

* 前向传播和反向传播：216毫秒（ms）
* 16条PCIe通道CPU-> GPU传输：大约2毫秒（理论上为1.1毫秒）
* 8条PCIe通道CPU-> GPU传输：大约5毫秒（2.3毫秒）
* 4条PCIe通道CPU-> GPU传输：大约9毫秒（4.5毫秒）

因此，PCIe通道从4个增加到16个将使性能提高大约3.2％。当然，如果将[PyTorch的数据加载器](https://pytorch.org/docs/stable/data.html#torch.utils.data.DataLoader)与页锁定一起使用，则性能将增加0％。因此，如果您使用单个GPU，请不要在PCIe通道上浪费金钱！

选择CPU PCIe通道和主板PCIe通道时，请确保选择支持所需的GPU​​数量。如果您购买支持2个GPU的主板，并且最终希望拥有2个GPU，请确保购买的CPU支持2个GPU，但不一定要查看PCIe通道。

## PCIe通道和多GPU并行

在数据并行的多GPU上训练网络，PCIe通道是否重要？我已经[在ICLR2016上发表了一篇关于此的论文](https://arxiv.org/abs/1511.04561)，我可以告诉您，如果您有96个GPU，那么PCIe通道就非常重要。但是，如果您拥有4个或更少的GPU，则无关紧要。如果您在2-3个GPU之间做并行化，那么完全不用关心PCIe通道。对于4个GPU，应确保每个GPU可以支持8个PCIe通道（总共32个PCIe通道）。由于几乎没有人运行带有超过4个GPU的系统，这是一条经验法则：不要花额外的钱为每个GPU获得更多的PCIe通道, 这无关紧要！

## 所需的CPU核数

为了能够明智地选择CPU，我们首先需要了解CPU及其与深度学习的关系。CPU对于深度学习有什么作用？在GPU上运行深层网络时，CPU很少进行计算。通常，它（1）启动GPU函数调用，（2）执行CPU函数。

到目前为止，CPU上做的最多的是数据预处理。有两种不同的数据处理策略，它们具有不同的CPU需求。

第一个策略是在训练时预处理：

循环执行以下三步：

1. 加载小批量数据
2. 预处理小批量数据
3. 小批量训练

第二种策略是在进行任何训练之前进行预处理：

1. 预处理数据
2. 循环处理以下两步：
   1. 加载预处理的小批量数据
   2. 小批量训练

对于第一种策略，具有多个内核的CPU可以显着提高性能。对于第二种策略，您不需要非常好的CPU。对于第一种策略，我建议每个GPU至少有4个线程, 通常每个GPU两个内核。我尚未对此进行严格测试，但是内核/GPU每增加1，性能大约提升0-5％。

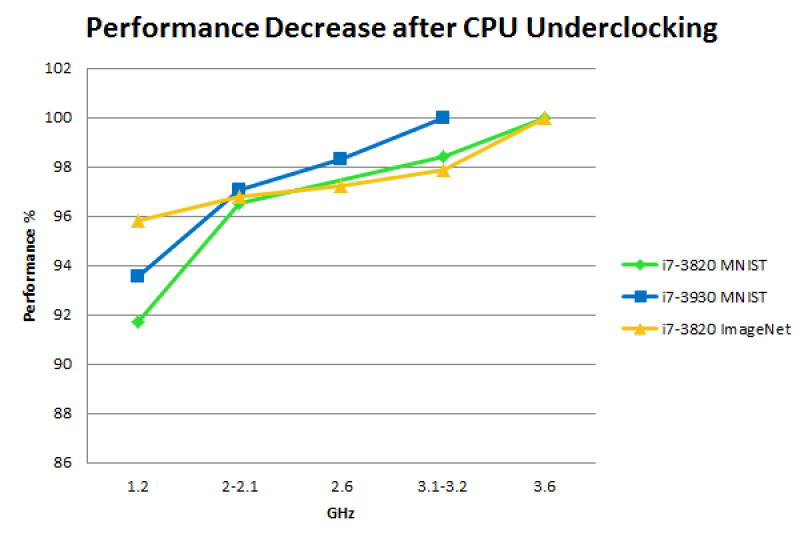
对于第二种策略，我建议每个GPU至少有2个线程-通常每个GPU一个内核。如果您使用第二种策略，则拥有更多核心时，性能不会显着提高。

## 所需的CPU主频

人们在考虑快速CPU时，通常通常会先考虑主频。4GHz比3.5GHz好吗？比较具有相同架构的处理器（例如“ Ivy Bridge”）通常是正确的，但是在处理器之间的比较却不佳。同样，它也不总是最佳的性能衡量标准。

在深度学习的情况下，CPU几乎不需要执行任何计算：要么增加一些变量，要么评估一些布尔表达式，在GPU或程序内进行一些函数调用, 所有这些都取决于CPU内核主频。

尽管这种推理似乎很合理，但事实是，当我运行深度学习程序时，CPU的使用率为100％，那么这是什么问题呢？我做了一些CPU降频实验以找出答案。

[](https://i0.wp.com/timdettmers.com/wp-content/uploads/2015/03/cpu_underclocking2.png)

**MNIST和ImageNet上的CPU降频**：性能以, 在不同CPU主频下, 200轮MNIST或1/4轮ImageNet所花费的时间, 来衡量. 以最大主频作为每个CPU的基线。作为比较：从GTX 680升级到GTX Titan的性能约为+15％；从GTX Titan到GTX 980还有+ 20％的性能；GPU超频可为任何GPU带来约5％的性能

注意，这些实验是在过时的硬件上进行的，但是，对于现代CPU/GPU，这些结果仍应相同。

# 硬盘/ SSD

硬盘通常不是深度学习的瓶颈。但是如果用法不对, 也会造成麻烦：如果在需要时从磁盘读取数据（阻塞模式），则32位的ImageNet小批量数据, 在100 MB/s硬盘驱动器上, 将花费您大约185毫秒！但是，如果采用异步加载（例如，torch vision loader）, 则可以在神经网计算时用185毫秒加载小批量数据，大多数基于ImageNet的深度神经网络的计算时间约为200毫秒. 由于在计算当前小批量时加载下一个小批量，因此不会有任何性能损失。

但是，我建议使用SSD来提高舒适度和生产率：程序启动和响应速度更快，并且处理大型文件的速度要快得多。如果您购买NVMe SSD，则与常规SSD相比，您将获得更加流畅的体验。(M.2接口的NVMe的顶级配置可以达到3GB/s的读取速度, 太惊人了)

因此，理想的配置是: 一个用于数据集的大而慢的硬盘驱动器, 以及一个用于提高生产率和舒适度的SSD。

# 电源装置（PSU）

PSU最好足够强大, 以便支持将来的配置。GPU的能效通常会随着时间而提高。因此，尽管需要更换其他组件，但PSU应该可以使用很长时间，因此，多花点钱在PSU上是很划算的。

通常可以将CPU和GPU的瓦数相加, 再加10%作为其他组件的瓦数并用作功率峰值的缓冲区, 来计算所需的瓦数。例如，如果您有4个GPU，每个功耗为250瓦，一个CPU的功耗为150瓦，那么您将需要一个最小为4×250+150+100=1250瓦的PSU。为了确保可靠，我通常会再添加10％，这样总和将达到1375瓦。四舍五入配一个1400瓦PSU。

一个需要注意的重点是，即使PSU具有所需的功率，它也可能没有足够的PCIe 8针或6针连接器。确保PSU上有足够的连接器以支持所有GPU！

另一个重点是购买具有高能率等级的PSU, 尤其是如果您有许多GPU并且要运行很长的时间。

在满功率（1000-1500瓦）上运行4 GPU系统训练一个卷积网两周的时间将达到300-500 kWh，在德国电价高达20分/每度电, 这意味着将达到60- 100欧元（66-111美元）。如果这个是效能为100％时的价格, 那么用能效为80％的电源训练这样的网络将使成本额外增加18-26€. 夸张吧?! 对于单个GPU而言，这要少得多，但要点仍然成立, 在高能效电源上多花点钱是很有意义的。

全天候使用多个GPU将大大增加您的碳排放，甚至这比你的交通（譬如飞机）和其他因素还要多。如果你想要负责的话，可以用[纽约大学机器学习语言组](https://wp.nyu.edu/ml2/carbon-neutral-lab/)（ML2）的做法, 很容易办到，也很便宜，可以作为深度学习研究者的标准做法。

# CPU和GPU散热

散热很重要，并且可能是一个严重的瓶颈，与较差的硬件选择相比，散热会大大降低性能。对于CPU，可以使用标准的散热器或多合一（AIO）水冷却解决方案，但是对于GPU来说，您需要特别考虑。

## 风冷

对于单个GPU或如果多个GPU之间有间隔（在3-4个GPU的情况下为2个GPU），风冷是安全可靠的。但是，当您尝试冷却3-4个GPU时，就很可能铸下大错，在这种情况下，需要仔细考虑。

现代GPU在运行算法时会提高速度，从而将功耗提高到最大，但是一旦GPU遇到温度限制（通常为80°C），GPU就会降低速度，从而使温度不要过高。这可避免GPU过热而损坏, 同时实现最佳性能。

但是，常规的风扇速度预编程时间表不适合深度学习程序，因此在开始深度学习后几秒钟内, 显卡温度就达到了温度阈值。结果性能会降低（0-10％），对于多个GPU扎堆发热的情况, 这可能更严重（10-25％）。

由于NVIDIA GPU是最重要的游戏GPU，因此它们已针对Windows进行了优化。您可以在Windows中单击几下来更改风扇调度，但在Linux中则不能这样，并且由于大多数深度学习库都是为Linux编写的，因此这是一个问题。

在Linux下, 唯一办法是修改Xorg服务器（Ubuntu）的配置，修改其中的“ coolbits”选项。对单GPU来说这很有，但是如果是多GPU，因为有些是无头的 (没有连接监视器) ，需要模拟一个显示器, 这比较怪异又难搞。我尝试了很长时间, 被折磨了好几个小时, 并尝试使用实时启动CD来恢复我的图形设置，但是一直没法在无头GPU上设定好这个配置。

对3-4个GPU使用风冷，最重要的考虑因素是注意风扇的设计。鼓风机式风扇将空气排出到机箱后部，从而将新鲜，凉爽的空气吹入GPU。开放式风扇则是带动GPU周边的空气来冷却GPU, 但是热气还是在机箱里面转，如果有多个GPU彼此相邻，那么周围就不会有冷空气，并且这种用开放式风扇的GPU会越来越热，直到它们自行降低温度以达到更低的温度为止。绝对不要在3-4个GPU配置中使用开放式风扇。

## 适用于多个GPU的水冷方案

另一个更昂贵更精巧的选择是使用水冷。如果您只有一个GPU，或者两个GPU之间（在3-4个GPU板上有2个GPU）之间有空间，我不建议您使用水冷。但是，水冷可确保即使是最强大的GPU在4 GPU设置中也能保持凉爽，而这是风冷没法做到的。水冷的另一个优点是它的运行安静得多，如果您在其他人工作的区域中运行多个GPU，这将是一大优势。每个GPU的水冷成本约为100美元，还需要支付一些额外的前期成本（约50美元）。水冷还需要花费额外的精力来组装计算机，但是有许多详细的指南，总共只需要花费几个小时。维护不应那么复杂或费力。

## 大机箱散热？

我为深度学习集群购买过塔式机箱，因为它们为GPU区域提供了额外的风扇，但是我发现这基本上是无关紧要的：降低大约2-5°C，不值得投资，而且外壳大笨重。重点实际上是直接在GPU上制冷的方案, 不要为了GPU的冷却能力选择昂贵机箱。节约点. 机箱能满足GPU的基本要求就够了, 别多折腾.

## 制冷总结

简单来说：对于单个GPU来说，空气冷却是最好的。对于多个GPU，您应该使用鼓风机式的风冷方案并接受微小的性能损失（10-15％），或者多花点钱搞个水冷，水冷配置起来麻烦点但是没有性能损失。在某些情况下，空冷和水冷都是合理的选择。但是，我通常建议使用空气冷却来简化操作-如果运行多个GPU，请使用鼓风机式GPU。如果你想使用水冷方案, 尽量找一体化 (all-in-one) 解决方案.

# 主板

您的主板应具有足够的PCIe端口以支持您要运行的GPU数量（即使您有更多PCIe插槽，通常也限于四个GPU）；请记住，大多数GPU的宽度为两个PCIe插槽，因此，如果要使用多个GPU，请购买一块在PCIe插槽之间有足够空间的主板。确保您的主板不仅具有PCIe插槽，而且实际上支持您要运行的GPU设置。如果您在newegg上搜索所选择的主板并查看规格页面上的PCIe部分，通常可以在其中找到信息。

# 电脑机箱

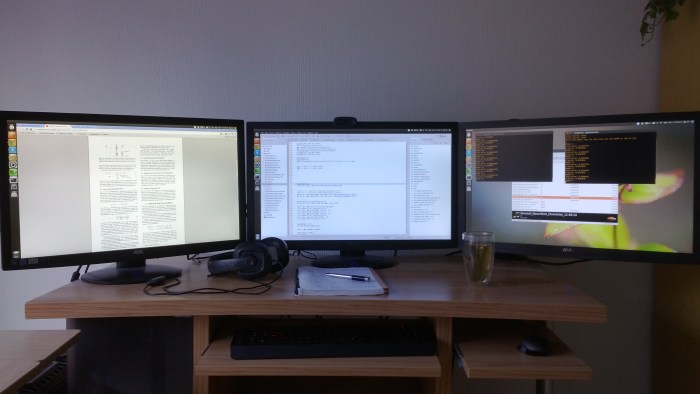
选择机箱时，奥确保把全长尺寸的GPU插在主板上还能放得下。大多数机箱都支持全长GPU，但如果是小型机箱下，就要确认下。检查尺寸和规格；还可以尝试在Google上搜索该款型的图片，看看是否找到带有GPU的图片。

如果使用定制的水冷系统，请确保机箱有足够的空间容纳散热器。如果将水冷用于GPU，更加要注意这一点。每个GPU的散热器都需要一些空间, 确保GPU能放得进去。

# 显示器

开始的时候我觉得写显示器有点傻. 但是其中差别真的很大，而且非常重要，我不得不写写它们。

我在3台27英寸显示器上花费的钱可能是我有史以来花的最值的钱。使用多台显示器时，生产率大大提高。如果必须使用一台显示器，我会感到极大的限制。不要在这件事上限制自己。如果您无法有效地操作一个快速深度学习系统，那么它有什么用？

[](https://i2.wp.com/timdettmers.com/wp-content/uploads/2015/03/2015-03-04-13-58-10.jpg)**进行深度学习时的典型显示器布局：**左：论文，Google搜索，gmail，stackoverflow；中：代码；右：输出窗口，R，文件夹，系统监视器，GPU监视器，待办事项列表和其他小型应用程序。

# 关于组装PC的一些话

许多人不敢组装计算机。硬件配件很贵，大家都不想犯错。这很简单，因为不搭的配件就是无法配合在一起。主板手册通常非常具体地说明如何组装所有东西，如果您没有经验，则有大量的指南和分步视频可以指导您完成整个过程。

组装计算机的妙处在于，一旦完成计算机，您就知道组装计算机的所有知识，因为所有计算机都是以相同的方式组装的, 所以组装计算机将成为您的一项生活技能, 以后将会一次又一次地用到。所以没有理由退缩！

# 结论

**GPU**：RTX 2070或RTX 2080 Ti。GTX 1070，GTX 1080，GTX 1070 Ti和GTX 1080 Ti也都不错！  
**CPU：**

* 每个GPU 1-2个内核，具体取决于您预处理数据的方式。
* > 2GHz；CPU应该支持您要运行的GPU数量。
* PCIe通道数无关紧要。

**内存**：

* 时钟速率(主频)无关紧要
* 购买最便宜的内存。
* 至少购买尽可能多的CPU 内存以匹配最大的显存。
* 仅在需要时购买更多内存。
* 如果您经常使用大型数据集，则更多的内存可能很有用。

**硬盘驱动器/ SSD**：

* 用于数据的硬盘驱动器（>= 3TB）
* 将SSD用于舒适性和预处理小型数据集。

**PSU**：

* 加上GPU和CPU的瓦数。然后将总数乘以所需瓦数的110％。
* 如果使用多个GPU，使用能效等级的电源。
* 确保PSU有足够的PCIe连接器（6 + 8针）

**冷却**：

* CPU：获得标准的CPU冷却器或一体式（AIO）水冷却解决方案
* GPU：
* 使用空气冷却
* 如果购买多个GPU，选配备鼓风机式风扇的GPU
* 在Xorg中设置coolbits标志配置以控制风扇速度

**主板**：

* 获得（未来）GPU所需的PCIe插槽数量（一个GPU占用两个插槽；每个系统最多4个GPU）

**监视器**：

* 附加的监视器可能比附加的GPU使您的工作效率更高。

2018年12月14日更新：重新整理了整个博客文章，并提供了最新建议。  
2015年4月22日更新：删除了对GTX 580的建议