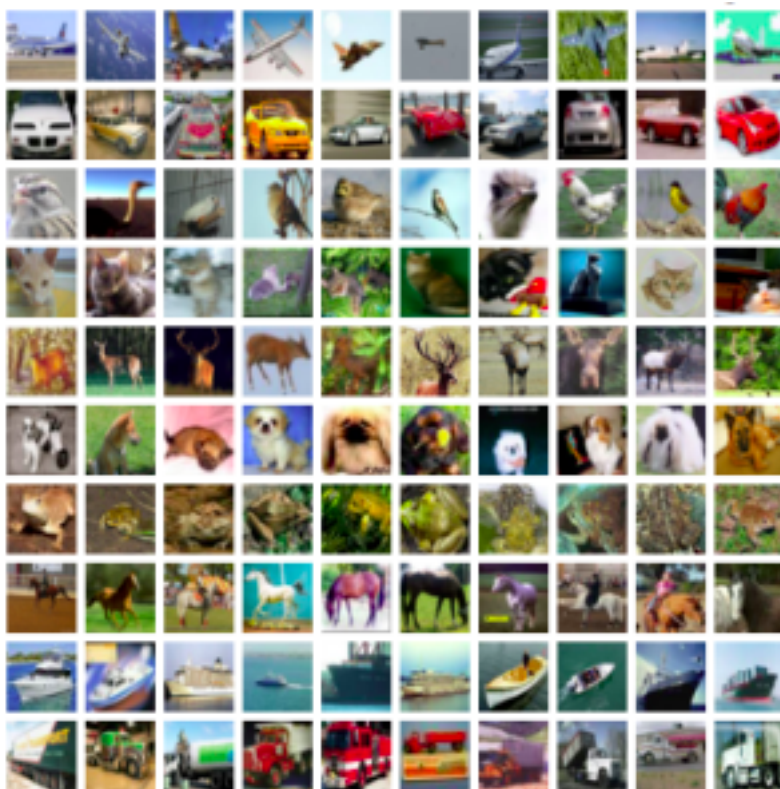


卷积网络训练太慢？Yann LeCun：已解决CIFAR-10，目标 ImageNet

摘要：CIFAR-10竞赛之后，卷积网络之父Yann LeCun接受相关采访。他认为：卷积网络需要大数据和高性能计算机的支持；深层卷积网络的训练时间不是问题，运行时间才是关键。Yann LeCun还分享了他正在做的一些最新研究。

Kaggle近期举办了一场 关于CIFAR-10数据集的竞赛，该数据集包含有6万个32*32的彩色图像，共分为10种类型，由 [Alex Krizhevsky](#), [Vinod Nair](#)和 [Geoffrey Hinton](#)收集而来。

很多竞赛选手使用了卷积网络来完成这场竞赛，其中一些在该分类任务中靠着超乎人类能力的表现而得分。在本系列的博客中，我们将会分别采访三位选手和卷积网络之父、Facebook人工智能实验室主任和纽约大学教授 [Yann LeCun](#)。



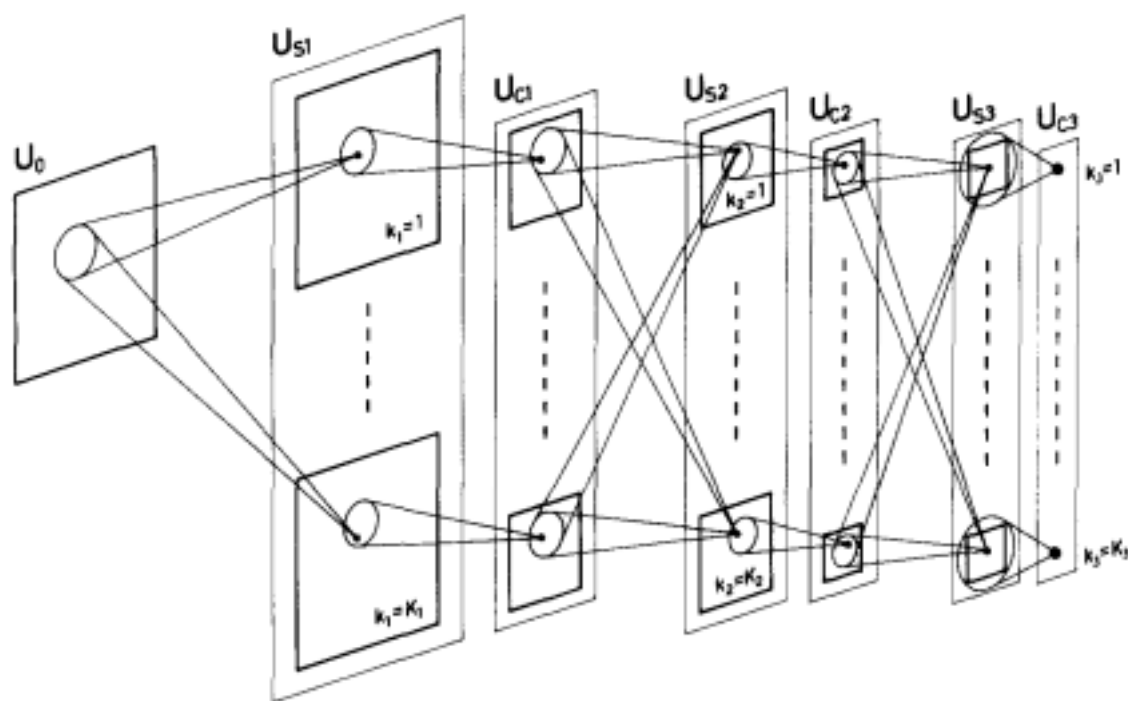
上图是cifar-10数据集的样例

以下是Yann LeCun的访谈内容：

还有哪些科学家为卷积网络的成功作出了巨大贡献？

毫无疑问，日本学者 [Kunihiko Fukushima](#) 提出的神经认知机

(NeoCognitron) 有着启迪性的意义。尽管卷积网络 (convnets) 的早期形态并未包含NeoCognitron太多的影子，但是我们使用的版本（有pooling层的）却深受其影响。



示意图演示的是神经认知机中层与层之间的相互连接。Fukushima K.

（1980年）在 [神经认知机](#) 一文中写道：模式识别机制的自组织神经网络模型不受位置变化的影响。

能回忆一下早期在卷积网络研究时出现的“顿悟”时刻或者突破性进展吗？

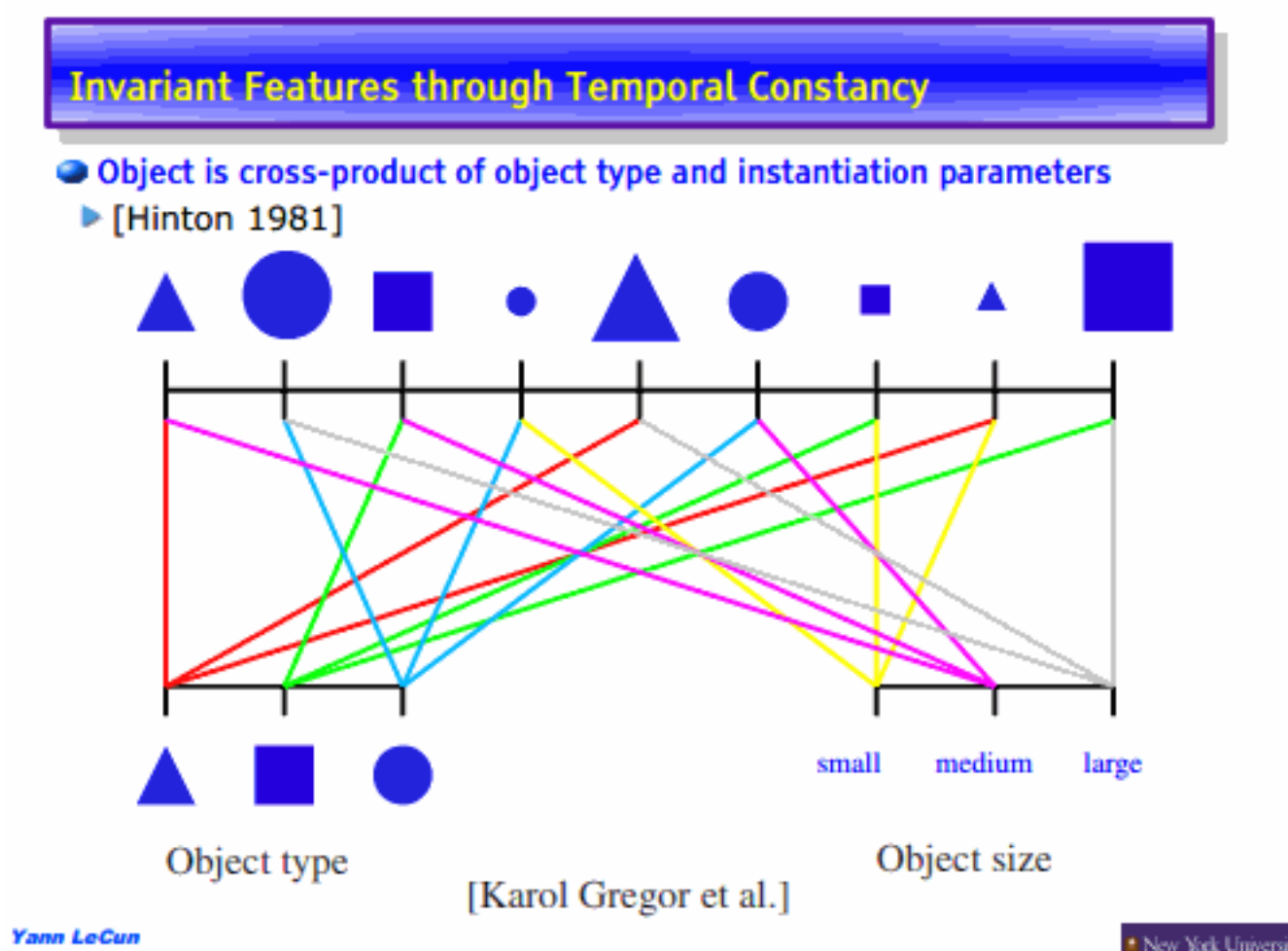
从1982年左右，我一直就本地节点的多层网络进行研究（尽管缺乏正确的学习算法，那时候 [Backprop](#) 即BP算法还没出现）。在1988年的时候我拿到了博士后学位，那时候我在做共享权值网络（shared weight nets）的研究试验。

没能早一点进行卷积网络相关研究的原因相当简单：缺乏软件和数据。一到贝尔实验室之后，我接触到了大量数据集和运行飞快的电脑（在当时算是），因而可以试着构建完整的卷积网络，而且令人惊喜的是其运行良好（尽管需要两周的训练）。

您对卷积网络近期在目标识别方面大热的看法如何？是否曾预料过？

是的，我知道肯定会这样，只是时间问题罢了，取决于数据集是否足够大，计算机是否足够强劲，能够支持深度学习算法在设计视觉系统方面比人类工程师做得更加优秀。

麻省理工学院2011年8月有一个“计算机视觉前沿”研讨会，我在上面 [演讲](#) 的题目为《五年内所有人都会学习该特性（你不妨现在就开始）》，[David Lowe](#)（[SIFT](#) 算法的发明者）也表示了同样的看法。



LeCun Y.在2011年演讲时的一张幻灯片

但是我仍然为变更发生的速度，还有卷积网络较其他手段的优秀程度感到惊诧，在我的预想中，过渡会更平缓一些。同样，我曾寄望于“无监督学习”的优秀表现。

AT&T的特性认知模型不止是一个简单的分类器，而是完整的管道。能否深入介绍您的团队所面临的实现问题？

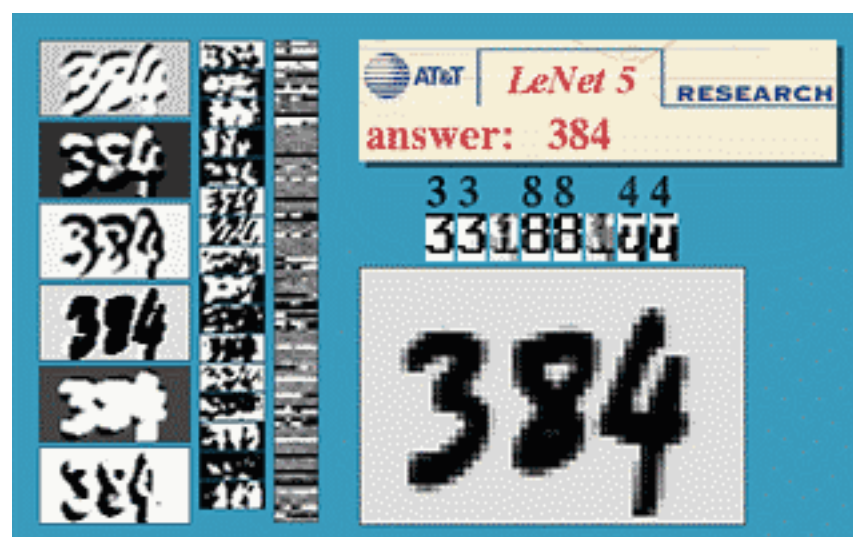
为了做到这一点，我们必须实现自己的程序语言同时编写自己的编译器。早在1987/1988年，我和[Leon Bottou](#)一同编写了一个名叫SN的神经网络模拟器，它是一个包含numerical library的LISP解译器（多维数组，神经网络图形……）。我们在贝尔实验室中使用它开发出了第一个卷积网络。

在90年代早期，我们想要在产品中使用自己的代码。最初我们雇佣了一支开发团队，将我们的Lisp代码转化成C/C++，但是运算结果的系统无法轻易得到改善（它并非一个优秀的研发平台），因此我和Leon还有[Patrice Simard](#)一起为SN编写了一个编译器，用来开发下一代OCR引擎。

该系统率先将节点分段、卷积网络和图解模型集成起来，点对点地进行整体训练。

图解模型被叫做“图变换网络”，从概念上来说与条件随机场（CRF）或者结构化的感知机（先于其产生）类似，但是允许包含非线性评分函数（CRF和结构化的感知机都只允许线性函数）。

该系统的整体基础架构以SN编写和编译，1996年时可部署在自动化机器上并读取图形字符，90年代末期可读出10%到20%的图形字符。



AT&T LeNet 5动画演示

与其他方式比较，卷积网络的训练十分缓慢。您如何在实验与增加的模型训练时间之间进行取舍？典型的迭代开发是什么样的？

在实验中，最佳的大型学习系统训练总得花费两到三周，这还没算任务、方法、硬件或者数据。

我不知道卷积网络是否“太慢”，你是说与什么相比呢？可能训练上比较慢，但是如果不这样的话，也需要工程师数月徒劳的努力。另外，（在训练后）卷积网络运行起来实际上是很快的。

在一个实际的应用中，没人关心训练到底要花多长时间，人们只关心运行需要多久。

最近关于卷积网络哪些论文最让您兴奋？哪些是我们应当留意的？

在过去的20多年中，有许许多多关于卷积网络和深度学习的想法都默默无闻，没人关心那些，发表论文总是十分艰难，因此有很多想法并未经过尝试

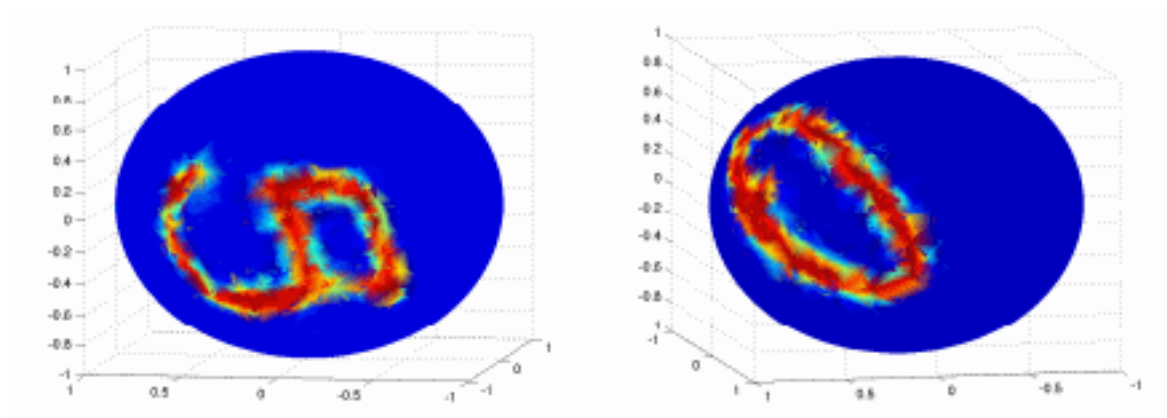
或发表，再或者就是尝试发表了但却被完全忽视并迅速忘却了，谁还记得基于人脸监测的首个有效尝试使用的是卷积网络（早在1993年，比 [Viola-Jones](#) 还要早八年）？



Vaillant R., Monrocq C., LeCun Y. 《肖像中目标定位的原始手段》，1993年

今天，看到这么多前途光明的年轻人在这个主题上如此积极投入并提出这么多新想法和新应用，真是让人惊喜不已。硬件/软件基础架构越来越好，在几个小时或者几天里进行大型网络的训练逐渐变为可能，因此人们可以进行更多想法的尝试。

我十分感兴趣的一个idea是“光谱卷积网络”。这是在ICLR 2014上我在纽约大学实验室的同侪发布的一篇论文，课题是广义化的卷积网络可以适用于任意图像（规则的卷积网络可以适用于1D，2D或3D数组，按图像可以被看成规则网格）。的确还存在一些实际的问题，但是它开启了一扇大门，让我们看到了通往卷积网络非结构化数据的更多应用方向。



球体内的MNIST数字

出处：Bruna J., Zaremba W., Szlam A., LeCun Y. 《光谱网络与图形方面的深度本地连接网络》，2013年

对于卷积网络还有周期性网络在自然语言理解方面的应用，我倍感兴味（跟着 [Collobert](#) 和 [Weston](#) 开创性的工作）。

鉴于人类的错误率被估算在6%左右，而Graham博士展示的结果在4.47%，你是否认为CIFAR-10已经被解决了？

这个问题同 [MNIST](#) 一样已经得到解决，但坦率来讲，相比CIFAR-10现下人们对于 [ImageNet](#)（图像识别目前最大的数据库）更感兴趣。在这层意义上，CIFAR-10并非是一个“真正”的问题，但是对于一个新算法来说这不是一个糟糕的基准。

工业界对卷积网络更加广泛的采用还需要什么条件？卷积网络的训练和建立所需的软件会变得更加简单吗？

要看你在说什么，现在卷积网络在工业界已经无处不在（或者近似无处不在），包括Facebook, Google, Microsoft, IBM, Baidu, NEC, Twitter, Yahoo!等等。

即便如此，事实上以上所有公司都有重要的研发资源，而卷积网络训练对于小公司或者科技没那么先进的公司来说还是很有挑战性的。

如果你没有接受过训练的话，接受卷积网络的训练仍然需要相当的经验 and 时间投入，但是很快会有几个简单的高效后端开源包出现。

我们离卷积网络的极限还有多远？或者说CIFAR-100会被“解决”吗？

我不觉得这是个好测试，[ImageNet](#) 会更好。

通过浅层网络的训练可以执行类似复杂的、设计良好的、更深的卷积架构。深度学习网络真的需要那么深么？

没错，深度学习网络需要如此。尝试通过浅层网络训练来模拟在ImageNet上训练过的深层卷积网络，你会发现，理论上深度学习网络可以与浅层的相接近，但是在复杂的任务上，浅层网络相距甚远。

您大部分的学术著作在本质上都有高度实用性，是故意为之还是公司的特意要求？是否能够告诉我们理论与实践之间的差别？

从2003年开始我就开始从事学术工作，同时还兼职纽约大学的教授。理论研究同时也帮助我理解事物。理论通常会帮助我们理解什么是可能的，什么是不可能的，为我们指明最合适的办事方式。

但是有时候理论会对思维产生限制，某些人不会使用一些理论模型，因为相关的理论太过晦涩，但是通常来说，一个在人们得知为什么它会运行良好的原因之前就已经运行良好的技术，理论上会获得更全面的理解。

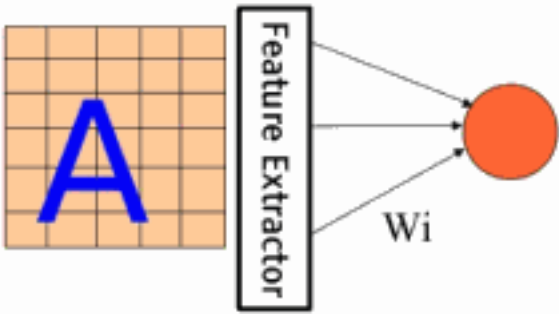

一个理论上完全获得理解的东西会反而让你受到束缚，你会先入为主地使用简单的方式。

另外，有时候理论会让我们盲目，比如一些人会因为核心方法附属的cute math而眼花缭乱，但是就像我之前说过的那样，最终的核心机器是那些执行“美化模版匹配”的浅层网络。这点完全没错（SVM是很好的方式），但是它有着可怕的局限性，我们应当充分警惕。

Y LeCun

► This Basic Model has not evolved much since the 50's

- The first learning machine: the **Perceptron**
 - Built at Cornell in 1960
- The Perceptron was a **linear classifier** on top of a simple **feature extractor**
- The vast majority of practical applications of ML today use glorified **linear classifiers** or glorified template matching.
- Designing a feature extractor requires considerable efforts by experts.


$$y = \text{sign} \left(\sum_{i=1}^N W_i F_i(X) + b \right)$$


LeCun Y.的幻灯片，从稳定特征来学习层次结构，2013年

在您看来，一个运作良好且不受“为什么它会运作良好”这一理论原因影响的卷积网络是什么样的？通常您是否偏好执行更甚于理论？如何平衡？

我不觉得执行与理论之间需要抉择，如果能执行，就会有相关可解释的理

论。

另外，你谈及的是什么样的理论？是说泛化界限吗（[generalization bound](#)）？卷积网络有着限定的VC维（VC Dimension），因此两者一致并有着典型的VC维。你还想要什么？是像SVM中那样更严格的界限吗？就我所知，没有一个理论的界限会足够严格，可以适用实际需求，所以我的确不太明白这个问题。当然，普通的VC维不够严格，但是非泛型范围（像SVMs）只会稍微欠缺一点严格性。

如果你希望的是收敛性证明（或保证），就有一点复杂了。多层网的损失函数不是凸值，因此简易证明该函数为凸不可行。但是我们都知实际上卷积网络几乎总是会收敛到同一水平，无论起点是什么（如果初始化完成正确）。有理论证明，有许许多多等效局部最小值，还有一个数字很小的“坏”局部最小值。因此收敛性并不会造成问题。

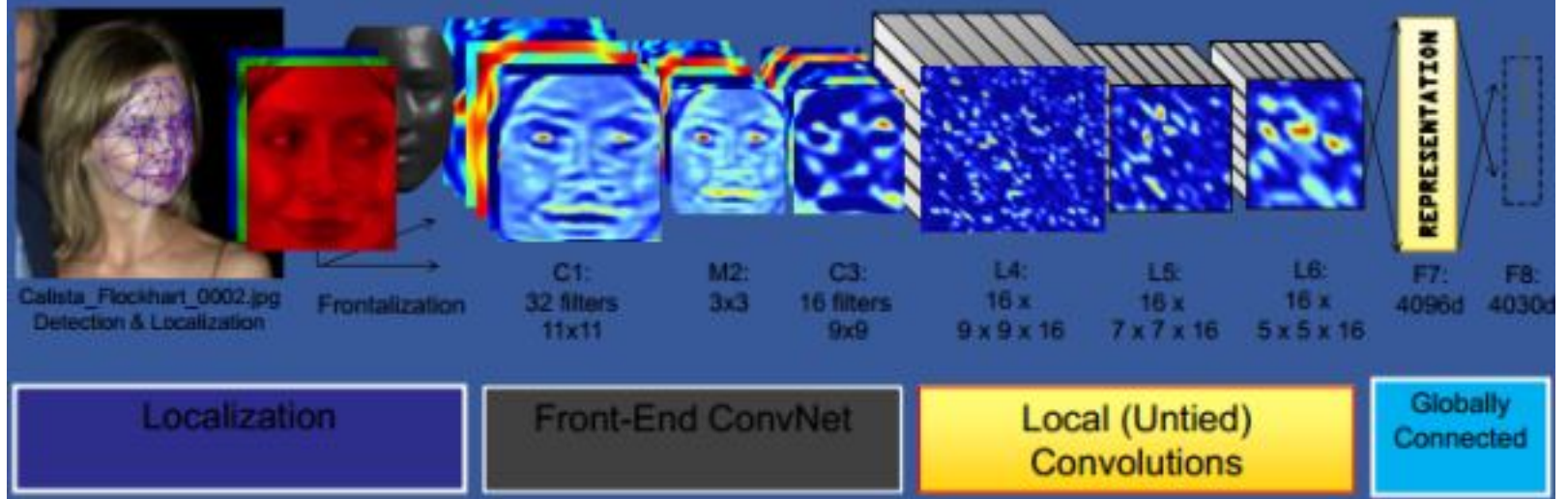
关于人工智能的大肆宣传您怎么看？您认为怎样的实践不利于这一领域（即一般的人工智能尤其是卷积网络）？

人工智能的大肆宣传非常危险，它令不同手段的人工智能相比过去减少了至少四倍，无论何时只要看到大肆宣传，不管它是来自报刊媒体、寻求投资的初创公司、寻找PR的大型公司还是寻找投资的学术人员，我都会大声喝止。

当然现在有相当一部分是关于深度学习的大肆宣传，我还没看到过专门关于卷积网络特别厉害的宣传，有更多是皮质层、神经元还有神经形态之类。与很多这类东西不同的是，实际上卷积网络在有用的任务上确实能带来好的结果，并在行业应用上广为应用。

Facebook有什么跟卷积网络相关的有趣项目吗？能谈一下基本情况吗？

人脸深层识别：人脸识别的卷积网络，同时也是图像标记的卷积网络，非常大。



描述体系结构的图形

出处：Taigman Y., Yang M., Ranzato M., Wolf L. 《不限制的人脸深层识别》，2014年

最近您发表了《四型严肃研究者》，您对自己怎么归类？

我是第3种，包含一点1和4的特质。

1. 想要在基本或理论层面解释或理解学习（也许是智能）的人。
2. 想要解决实际问题的人，对精神系统科学毫无兴趣。
3. 想要理解智能，建构智能机器，并且对大脑运作方式有部分兴趣的人。
4. 首要兴趣就是理解大脑的运作，但是认为需要先建构可运作的机器模型才能对其理解的人。

对CIFAR-10挑战的获胜者您有什么想说的吗？对于学习卷积网络的研究人员或是兴趣爱好者您有什么祝愿？对于CIFAR数据集或者问题呢？

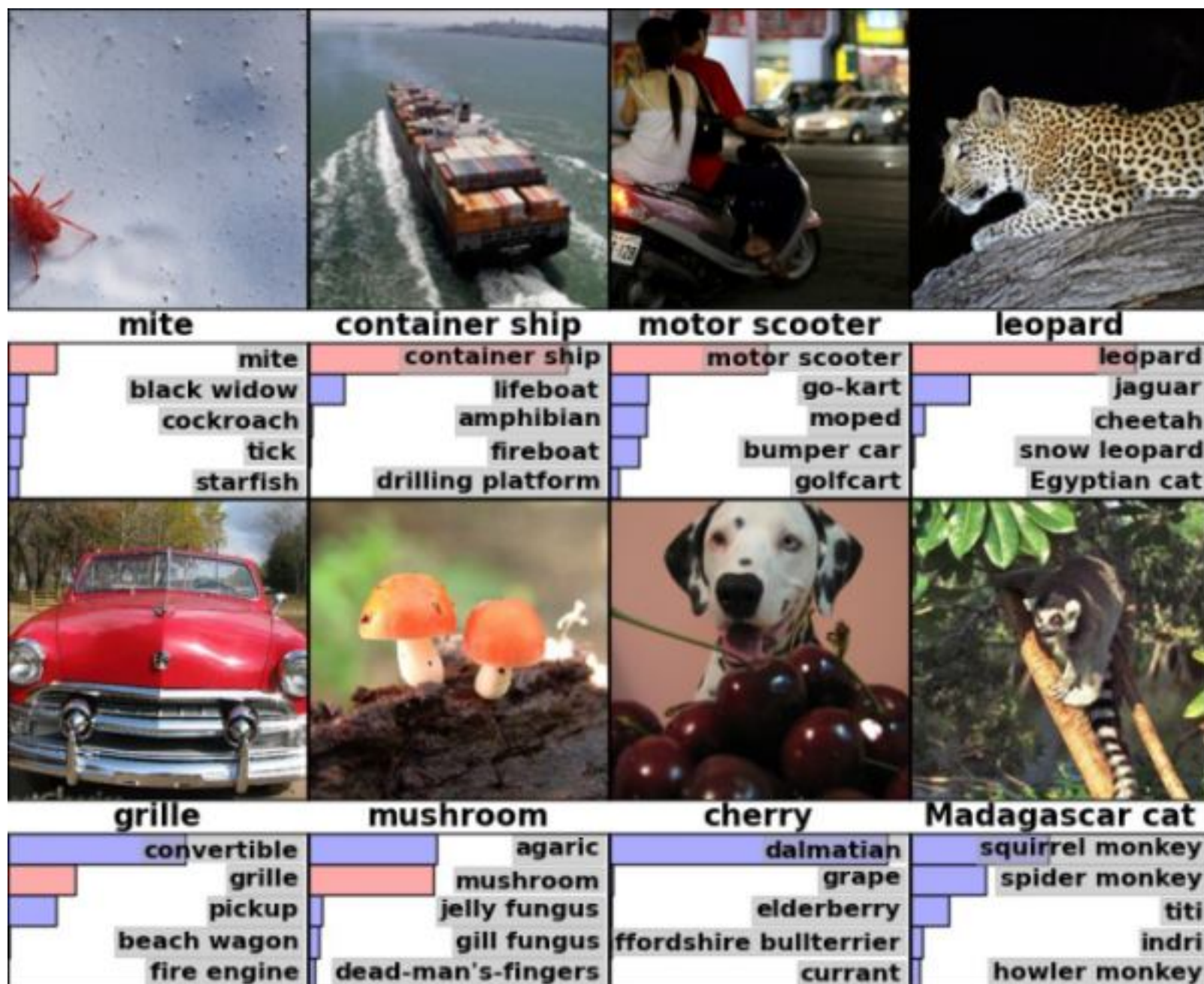
参与者的创造力和设计技巧给我留下了很深刻的印象，人们推动相关科技进步的行为很令人欣慰。

但是对于独立研究人员和兴趣爱好者来说，研究这些并将其应用于大型数据集变得越来越简单，我认为CIFAR-10的获胜者应当是ImageNet-1K-128*128，这会是具有1000种图像网络分类任务的版本，同时图像为标准化的128*128，我看到了一些优点：

1. 对于一个高端玩家的操作来说，网络够小，可以在合理时间内训练；
2. 你最终得到的网络可以在实践中为有用的应用所使用（像是机器人视

觉)；

3. 网络可以实时在嵌入式平台上运行，像是智能手机或 [NVIDIA Jetson TK1](#)。



ImageNet的预测

出处：Krizhevsky A., Sutskever I., Hinton. G.E. 《深层卷积神经网络的ImageNet分类》，2012年

拥有大量标记数据的需求可能会是个问题，对于无标记数据或者运用图片搜索引擎进行数据的自动标记化的网络训练上，您有什么看法？

有像是视频理解和自然语言理解这样的任务，我们计划会使用无监督学习方式。但是这些形式都有时间维度因素，会影响我们解决问题的手段。

明确来说，我们需要设计算法，这种算法可以学习感知世界的架构，并无需被告知所有事物的名称。我们中的很多人这些年甚至数十年来一直都在从事这方面的研究工作，但是都没有完美的解决方案。

您最新的研究是关于什么的？

就这个问题有两个答案：

1. 我个人从事的项目（份量足够作为项目的论文作者之一）
2. 正在做准备的项目，为其他人提供支持的项目，在概念层面提供建议的项目，在这些项目中我参与的程度并不足以作为项目的论文作者之一出现。

项目类型一大多在纽约大学，项目类型二大多在Facebook。

大体领域包括：

探索“不变量”特性的非监督学习，深层学习与结构化预测的结合，监督与非监督学习的一致性，解决学习长期依赖关系的问题，建构短期或暂存学习体系，学习计划与一系列的行动，优化功能的不同方式，从表象学习到综合推理（阅读Leon Bottou优秀的意见论文《从机器学习到机器推理》），学习对于高效执行推理的用处，以及众多其他课题。

原文链接：[Convolutional Nets and CIFAR-10: An Interview with Yann LeCun](#)（翻译/孙薇 责编/周建丁）