2/37 我们的project旨在对卷积神经网络算法进行加速。卷积神经网络是目前非常流行的、被广泛用于计算机视觉、图像分类、识别或处理等领域。

在有图形处理单元GPU的帮助下，我们在高效率的计算方面得到了不错的结果。

我们的接下来的展示，组织结构如下：

3/37 首先简单介绍卷积神经网络的结构、算法和应用；接着是CUDA和Caffe工具的安装；然后是我们做的算法测试实验，并针对性能和时间上给出了直观的比较结果。

5/37 卷积神经网络是一种特殊的人工神经网络，设计用于处理多维矩阵，比如彩色图像，由3个通道的二维矩阵像素值组成。其大致的结构如下，大致是一种串行的层级结构，输入图像像素，中间的像素称为特征图，最后产生网络的输出结果。

6/37 中间的层分别有卷积层、池化层，和在最后的全连接层。每一个特征图的单个元素都由前一层的部分元素通过与权值的乘积再求和得到。特征图越大，数量越多，所需要的权值数量也就越多，相应的计算量也就越大。

7/37 在上面的简单结构中，已经有成百上千的权重参数在网络中待学习，所以在训练过程中，需要非常大量的运算，特别是卷积。

实际上，我们是一个接着一个计算图中并列的几个特征图，这是编程过程中很通常的做法。但是如果我们把计算步骤并行化，即同时计算一层的多个特征图，因为他们之间的计算其实互不影响，那么训练过程就会大大改善。

8/37 CNN 已经在很多方面有了先进的应用。图像分类、目标检测、人脸识别和语义分割。如图所示，有大量这样的图片混合起来，用训练好的卷积神经网络模型可以识别出哪几张图片是飞机、鸟、汽车或猫；在分类的基础上，再检测出目标的位置，比如找到图中老虎的位置。

目前，这些研究正被学术机构，如纽约大学、蒙特利尔大学、多伦多大学、加州大学伯克利分校和斯坦福大学，带头进行，另外还有一些大公司，谷歌，facebook和微软的人工智能研究院也非常支持。