

合

新闻

体育

汽车

房产

旅游

教育

时尚

科技

财经

娱乐

更多

登录狐友

0

分享到

现代服务产业  
技术创新战略  
联盟1291 文章 37万  
总阅读

查看TA的文章&gt;

## 【Hinton碰撞LeCun】CNN有两大缺陷,要用capsule做下一代CNN

2017-08-20 20:40

人工智能 / 大脑



**【导读】**在本次演讲中, Hinton讨论了用“capsule”作为下一代CNN的理由。

他解释了“标准”的卷积神经网络有什么问题?结构的层次太少,只有神经元、神经网络层、整个神经网络。所以,我们需要把每一层的神经元组合起来,形成一个组,并装到“舱”(capsule)中去,这样一起来就能完成大量的内部计算,最终输出一个经过压缩的结果。“舱”(capsule)的灵感来自大脑皮层中的微柱体(mini-column)。CNN的代表人物是Yann LeCun,所以这也可以说成是两位大神在深度学习观点上的一次正面交锋。

### What is wrong with convolutional neural nets?

Geoffrey Hinton  
Department of Computer Science  
University of Toronto  
&  
Google Inc.

2017年8月17日, Hinton在加拿大多伦多菲尔兹研究所开讲,主题是《卷积神经网络有哪些问题》,这是加拿大新成立的“向量研究院”(Vector Institute)2017-2018机器学习的发展和应用课程的一部分。

2017年3月30日,Vector Institute宣布成立,Hinton是这一机构的首席科学顾问。发布会上Vector方面表示将致力于人工智能的前沿研究,专注在机器学习和深度学习领域的变革性研究。该研究院将与学术机构、孵化器、加速器、初创企业以及大公司展开合作,推动加拿大人工智能的研究及商业化应用。

卷积神经网络(Convolutional Neural Network, CNN)是一种前馈神经网络,它的人工神经元可以响应一部分覆盖范围内的周围单元,对于大型图像处理有出色表现。

卷积神经网络的集大成者是Yann LeCun,现Facebook人工智能实验室的主管,它被业界誉为“卷积神经网络之父”。在本次演讲中,Hinton也多次提到了LeCun的观点,提到两人在

大家都在搜: 传苹果iCloud主管离职

货到付款



热门图集



徐才厚、谷俊山的这些画面首次集中曝光 我一定要老实交代



小伙乌克兰开婚介 向国内“东欧美女”批发 林志玲穿透视裙展女王范 俯身秀事业线



太有才! 中秋的月亮都快被 玛丽莲梦露罕见孕照曝光 玩儿坏了

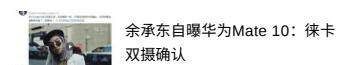


24小时热文

1 改变智能手机格局:谷歌Pixel 2 XL难堪大任

2 大厨师返乡开挖机 月薪比城市年薪多

3 苹果首款全面屏幕手机, iPhone X到底有什么不同?



KING NEWS Win10 Mobile 进入「维护期」, 卡西欧带来 2.5D 打印机 |...

合

新闻

体育

汽车

房产

旅游

教育

时尚

科技

财经

娱乐

更多

登录狐友

0

分享到



广告

## 搜狐号推荐



**IT之家**  
IT之家是业内领先的即时IT资讯和数码产品类网站。IT之家快速精选泛科技新闻...



**搜狐科技视界**  
搜狐科技官方原创账号。聚焦TMT领域大事件、大趋势和新变化，用我们的视角观...



**果壳网**  
面向都市科技青年们的社交网站。开放、多元的泛科技兴趣社区，并提供负责任...



**太平洋电脑网**  
喜欢数码的都在这儿！始于1999年的IT垂直媒体，值得信赖。



**柏铭007**  
关注科技，偶尔扯谈经济。

**超清 99678 米**  
晚上可以看（支持夜视）



广告

## What is wrong with "standard" neural nets?

- They have too few levels of structure:
  - Neurons, Layers, Whole Nets
- We need to group the neurons in each layer into "capsules" that do a lot of internal computation and then output a compact result.
  - A capsule is inspired by a mini-column.

联系我们

"标准"的卷积神经网络有什么问题？

结构的层次太少：神经元、神经网络层、整个神经网络

我们需要把每一层的神经元组合起来，形成一个组，并装到"舱"（capsule）中去，这样以来就能完成大量的内部计算，最终输出一个经过压缩的结果。

"舱"（capsule）的灵感来自大脑皮层中的微柱体（mini-column）。

## What does a capsule represent?

- Each capsule represents the presence and the instantiation parameters of a multi-dimensional entity of the type that the capsule detects.
- In the visual pathway, for example, a capsule detects a particular type of object or object-part.
- A capsule outputs two things:
  - 1. The probability that an object of that type is present.
  - 2. The generalized pose of the object which includes position, orientation, scale, deformation, velocity, color etc.

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

**0** 约是它所检测到的类型的一个多维实体的存在和实例化参数。

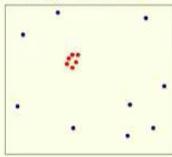
分享到 道上,一个“舱”会检测到物体的具体对象的类型。

一个“舱”会输出两个东西:

1. 被呈现的对象可能的分类;
2. 对象的大概状态,包括位置、朝向、大小、变形、体积和颜色等等。

**Capsules do coincidence filtering**

- A typical capsule receives multi-dimensional prediction vectors from capsules in the layer below and looks for a tight cluster of predictions.
- If it finds a tight cluster, it outputs:
  - 1. A high probability that an entity of its type exists in its domain.
  - 2. The center of gravity of the cluster, which is the generalized pose of that entity.
- This is very good at filtering out noise because high-dimensional coincidences do not happen by chance.
  - It's much better than normal "neurons"



“舱”可以完成同步过滤 (filtering)

一个典型的“舱”从下一层的“舱”中接收多维的预测向量,并且寻找一个更紧致的预测群 (cluster)。

如果找到了一个一个更紧致的预测群 (cluster), 它会输出:

一个高概率,即某一类型的实体存在在这个区间

群的引力中心,也就是实体的大概状态

这种方法在过滤噪音上做得非常好,因为高维度的一致性的发生并不是偶然。

它比一般的“神经元”表现得要好很多。

当下, LeCun和几乎所有人都在用的对象识别有什么问题?

**The current way to do object recognition**  
(Yann LeCun & now almost everyone else)

- Convnets use multiple layers of learned feature detectors. ☺
- In convnets, feature detectors are local and each type of detector is replicated across space. ☺
- in convnets, the spatial domains of the feature detectors get bigger in higher layers. ☺
- Feature extraction layers are interleaved with subsampling layers that pool the outputs of nearby feature detectors of the same type. ☺

当下用于对象识别的方法:

Convnets (卷积网) 使用多层学习到的特征检测器。(这一点很好)

在卷积网中, 特征的检测是局部的, 每一种类型的检测器被复制到整个空间中。(这一点很

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

0 次越高, 特征检测的空间领域变得越大。 (这一点很好)

分享到抽样层交叉存取, 将相同类型的相邻特征检测器的输出汇集到一起。 (这是

### The motivations for combining the outputs of replicated feature detectors

- 1. Pooling gives a small amount of translational invariance at each level.
  - The precise location of the most active feature detector is thrown away.
  - Maybe that's OK if the pools overlap a lot or if features code relative positions of other features (Ullman).
- 2. Pooling reduces the number of inputs to the next layer of feature extraction.
  - This allows us to have more types of feature at the next layer (with bigger domains).

将复制的特征检测器的输出进行结合的动机

池化在每一层都会给予一个小量的转换变量

最活跃的特征检测器的精确位置会丢失

可能, 这也是ok的, 如果池化堆叠很多次或者如果特征对其他特征的相对位置进行编码的话

池化减少输往特征提取下一层的数量

这将让我们在下一层拥有更多的特征类型 (更大的领域)

### What kind of a percept does a convnet have?

- The activations in the last hidden layer of a deep convnet are a percept.
  - The percept contains information about many of the objects in the image.
  - But what about the relationships between objects? It was not trained on relationships.
  - Let's just ask it (artificial cognitive science)
- Map the percept to the initial hidden state of a deep recurrent neural net and train the RNN to output captions (without retraining the convnet).

一个卷积网络中拥有什么类型的认知

深度卷积网络中最后一层的激活行为就是一个认知

感知包含了图像中许多物体的信息

但是, 物体之间的关系是怎样的? 关系的认知并没有经过训练

向一个深度循环神经网络的最初隐藏层的状态上加入上文提到的认知, 并且训练RNN来生成字幕 (不需要对卷积网络进行预训练)

反对池化的四点理由

## Four arguments against pooling

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

- **IT IS A BAD FIT TO THE PSYCHOLOGY OF SHAPE PERCEPTION.**
- 0 It is not explain why we assign intrinsic coordinate frames to objects and why they have such huge effects.
- 分享到 **It solves the wrong problem:** We want equivariance, not invariance. Disentangling rather than discarding.
- It fails to use the underlying linear structure:** It does not make use of the natural linear manifold that perfectly handles the largest source of variance in images.
- **Pooling is a poor way to do dynamic routing:** We need to route each part of the input to the neurons that know how to deal with it. Finding the best routing is equivalent to parsing the image.

对于池化（pooling），存在以上 4 点争论：

不合乎我们对形态感知的心理认知

它无法解释为什么把固有坐标系分派给对象后，会有如此明显的效果。

它在解决的是错误的问题

我们想要的是 equivariance，不是 invariance。想要的是 Disentangling，而不是 discarding。

它无法使用基本的线性结构

它不能利用能够完美处理图像中大量 variance 的自然线性流形。

池化对于做动态 routing 也很差

我们需要 route 进入神经网络的输入的每一部分，好知道如何处理它。找到最好的 routing 相当于为图像做 parsing。

### The tetrahedron puzzle: A demonstration of the power of coordinate frames

- Slice a solid tetrahedron into two pieces with a plane.
- How hard could it be to put the two pieces back together to make the tetrahedron?
- Most people take several minutes to solve it.
  - One MIT professor tried for 10 minutes and then wrote down a proof that it was impossible.
- How could a trivial task be so difficult?
  - We really need an explanation for this!

关于争论1：

四面体Puzzle：关于坐标系能做什么的演示

用一个平面将一个固体四面体切为两块

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

0 了10分钟,然后写下了个证明,证明这不可能实现

分享到 么这么难?我们需要一个解释。

## The inverse tetrahedron puzzle

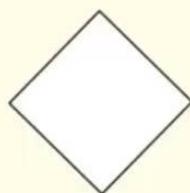
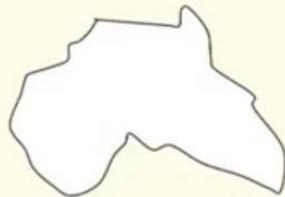
- Imagine slicing a solid tetrahedron with a plane so that you get a square cross-section.
- This is easy if you think about a tetrahedron in one way, and almost impossible if you think about it in the standard way.

反向四面体Puzzle

想象一下,用一个平面切开一个固体四面体,得到一个方形的cross-section;

如果你用一种方式去考虑这个四面体,做到这点并不难;而如果你用标准的方式去考虑这个四面体,就几乎不可能做到。

Some more psychological evidence that our visual systems impose coordinate frames in order to represent shapes (after Irvin Rock)



一些更多的心理学证据,显示了我们的视觉系统在抓住物体形状时,利用了坐标系。

## Argument 2: Equivariance vs Invariance

- Convolutional neural nets try to make the neural activities **invariant** to small changes in viewpoint by combining the activities within a pool.
  - This is the wrong goal.**
  - It is motivated by the fact that the final label**

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

**letter to aim for equivariance: Changes in point lead to corresponding changes in neural activities.**

分享到

**the perceptual system, it's the weights that code viewpoint-invariant knowledge, not the neural activities.**

## 关于争论2：Equivariance vs Invariance

卷积神经网络努力在让神经活动对视点上的小变化invariant，方法是通过在一个“池”内合并这些活动

—这个目标是错误的；

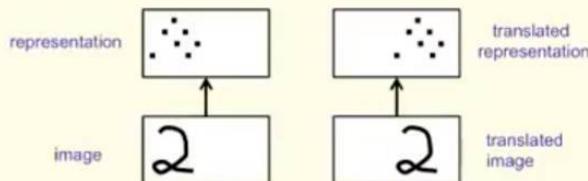
—它由这样一个事实驱动：最终的 label 需要 viewpoint-invariant

以equivariance为目标会更好：视点中的变化引发了神经活动中的相应变化

—在认知系统中，是 weights 编码了viewpoint-invariant knowledge，而不是神经活动。

## Equivariance

- Without the sub-sampling, convolutional neural nets give "place-coded" equivariance for discrete translations.



## Equivariance

没有sub-sampling，卷积神经网络为discrete translations 给出了“place-coded” equivariance。

## Two types of equivariance

- If a low-level part moves to a very different position it will be represented by a different capsule.  
– This is “place-coded” equivariance.
- If a part only moves a small distance it will be represented by the same capsule but the pose outputs of the capsule will change.  
– This is “rate-coded” equivariance.
- Higher-level capsules have bigger domains so low-level place-coded equivariance gets converted



新闻

体育

汽车

房产

旅游

教育

时尚

科技

财经

娱乐

更多

登录狐友

0

分享到ce

如果一个低级别部分移动到了一个非常不同的位置, 它会被不同的capsule表征

—这是“place-coded” equivariance。

如果一个部分仅移动了很短的距离, 它仍会被同样的capsule表征, 但capsule的输出将会变化

—这是“rate-coded” equivariance。

更高级别的 capsules 有更大的domain, 所以低级别的place-coded equivariance 转化为了高级别的 rate-coded equivariance。

### Argument 3: Extrapolating shape recognition to very different viewpoints

- Current neural net wisdom:
  - Learn different models for different viewpoints.
  - This requires a lot of training data.
- A much better approach:
  - The manifold of images of the same rigid shape is highly non-linear in the space of pixel intensities.
  - Transform to a space in which the manifold is globally linear (i.e. the graphics representation that uses explicit pose coordinates).
  - This allows massive extrapolation.

关于争论3：推算形状识别到非常不同的视点

目前的神经网络智慧

—学习用于不同视点的不同模型

—这要求大量训练数据

一个好得多的方法

—同样形状的图像流形在pixel intensities的空间里是高度非线性的

—向流形是全局线性的空间转化 (即, 图像表征使用了明确的形态坐标)

—这使得大量推算成为可能

### Generalize over viewpoints by using the globally linear manifold used by computer graphics

- Graphics programs use hierarchical models in which spatial structure is modeled by matrices that represent the transformation from a

合

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多

登录 狐友

coordinate frame embedded in each part.

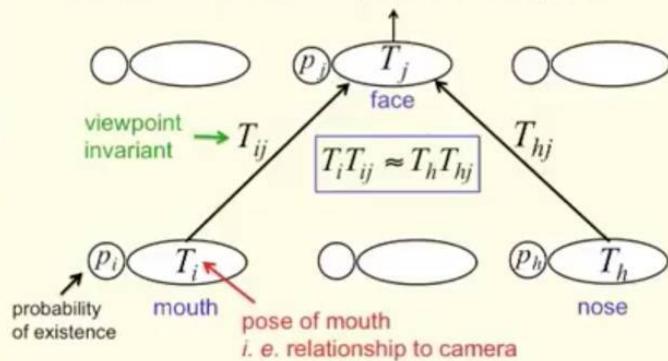
0 These matrices are totally viewpoint invariant.  
 This representation makes it easy to compute the relationship between a whole and the camera from the relationship between a part and the camera.

- Its just a matrix multiply!

使用计算机图像使用的全局线性流形在视点上泛化

### Two layers in a hierarchy of parts (coincidence filtering using the linear manifold)

- A higher level visual entity is present if several lower level visual entities agree on their predictions for its pose.



在部分层次中的两层（使用线性流形的coincidence filtering）

### Argument 4: We need to route information in the image to the neurons that can make sense of it.

- For complicated shapes, we do not want to replicate the knowledge across all locations with a small stride.
  - It's better to somehow route the information to a single capsule that can deal with it.
  - But the information could appear anywhere in the image.
- Eye movements are a great routing mechanism but they are slow and serial.
  - How else can neural nets route information?

关于争论4：我们需要 route 图像中的信息，好让神经网络可以理解

对于复杂形状来说，我们不希望用 small stride 复制所有位置的知识

—用某种方法能 route 信息到一个单独的能够处理它的 capsule会更好

—但这一信息可能会在图像中的任何位置出现



0 什么方法可以 route 信息？

分享到 [分享到微博](#)  
 所有内容均来自网络,我们对文中观点保持中立,对所包含内容的准确性,可靠性  
 提供任何明示或暗示的保证,请仅作参考。若有侵权,请联系删除。

文章来源：新智元

差一点

我们就擦肩而过了

有趣 有用 有态度

[返回搜狐，查看更多](#)

声明：本文由入驻搜狐号的作者撰写，除搜狐官方账号外，观点仅代表作者本人，不代表搜狐立场。

阅读 (685)

不感兴趣 投诉

## 本文相关推荐

- |                 |                     |                   |
|-----------------|---------------------|-------------------|
| cnn卷积神经网络       | cnn特征提取+tensorflow  | cnn中高维特征是什么       |
| cnn是如何提取图像特征的   | capsule+wardrobe+胶囊 | atos+capsule三维扫描仪 |
| time+capsule怎么拆 | cnn评一生必去50个旅行地      | cnn官网中文           |
| cnn双语新闻         | cnn评选中国最美旅行地        | cnn深度学习           |



¥368.00 ¥1598



¥378.00 ¥2986



¥368

客厅灯

床垫

冲锋衣

沙发垫

电脑广告

我来说两句

0人参与，0条评论

来说两句吧.....

[登录并发表](#)

搜狐“我来说两句”用户公约

还没有评论，快来抢沙发吧！

## 推荐阅读

[苹果iPhone X黑色版真机上手：未发售先偷跑](#)

IT之家 · 昨天 23:27

6

[失去薛之谦和鹿晗之后，是时候迎来不会崩坏的AI偶像了](#)

脑极体 · 昨天 21:43

5

[三星Note 9将用上屏下指纹：三家供应商疯抢订单](#)

PingWest品玩 · 今天 07:05

...

[押金退款难、拖欠工资 酷骑单车或10亿元变卖](#)

中国经济网 · 今天 07:25

3

[全新威驰遇见完美“型”影不离，钜惠来袭](#)

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多 [登录](#) [福友](#)

0 iPhone 8没有大火，但iPhone 7的火爆让国产手机竞争加剧

分享到 IT观察 · 今天 09:10

9

马斯克炮轰谷歌Clips人工智能相机：公然漠视隐私

IT之家 · 今天 09:14

...

鹿晗让微博服务器“葛优瘫”了，来看看这些年瘫痪过的服务器



北大新媒体 · 昨天 20:02

...

谷歌的营销套路：苹果做啥我做啥，然后靠AI技术“吊打”

猎云网 · 今天 09:14

...

散户必看 牛股大曝光



操盘手才知道的技巧，3天学会辨牛股，踩准买卖点

广告 · 今天 10:53

8点1氪 | 国庆档票房报收27.5亿创新高；中秋节当日微信红包收发量达...

36氪 · 今天 07:32

...

长假过后，反而更累了？你好好反思一下你是怎么放假的！



果壳网 · 昨天 22:46

11

鹿晗关晓彤公布恋情，微博尾巴暴露vivo和华为才是最大赢家



IT烽火台 · 今天 08:19

...

这只章鱼妈妈，创下了卵孵化时间的最长记录：四年半



果壳网 · 今天 09:02

...

这个微信每天推荐3只短线牛股，可以先观察几天看看效果！



广告 · 今天 10:53

地球如果变大，会发生什么？





新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多 [登录](#) [狐友](#)

0 俏有理论懂市场更受欢迎

01:52



分享到

nap的那些事(5): Instagram与Snapchat的相爱相杀



柳胖胖 · 昨天 23:59



从变形到磁场感知: 动物世界中的那些超能力



IT之家 · 今天 08:58

新东方雅思联手剑桥名师助您雅思轻松通过!



广告 · 今天 10:53

黑莓与TCL正式发布全触屏机型 BlackBerry Motion



PingWest品玩 · 今天 07:43

大家好, 这是鹿晗用来介绍女朋友的手机



3

虎嗅 虎嗅APP · 昨天 22:02



绕着月球画个心——NASA正式重启登月计划

NASA中文爱好者团队 · 昨天 23:27



互联网巨头集体“搅局”, 中国支柱行业将被彻底颠覆



36氪 · 今天 10:27

短线高手每天微信推荐3只短线牛股, 可以观察几天看看效果



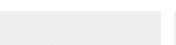
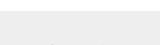
广告 · 今天 10:53

NASA的新研究显示月球可能有过大气层

煎蛋 · 今天 00:00



互联网巨头集体“搅局”, 中国支柱行业将被彻底颠覆



新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多 登录 狐友

0 1:27

分享到 科学家：时间旅行可能实现，并不是所有时钟都相同

IT之家 · 昨天 14:55 11

假期大事件 | 苹果正式调查iPhone8电池；法拉第创始团队仅剩贾跃亭等三人；亚马逊线下...

36氪 · 今天 08:36

加微信每天推荐3支短线牛股,可以先观察几天看看效果! 不花钱

广告 · 今天 11:7

绕着月球画个心——NASA正式重启登月计划

NASA中文爱好者团队 · 昨天 23:27

短视频变现，又一场“诗和远方”的美好骗局？

短视频工场 · 昨天 20:00

他16岁自学AI编程被谷歌关注了，你还在抱怨假期太少。。

AI星球 · 今天 09:01

【钛晨报】微信发布国庆长假数据报告：数百万人跨境旅游，香港成境外消费最多地区

钛媒体APP · 今天 06:57

ThinkPad 25周年限量版只是改回经典七行键盘的T470，你愿意为这情怀买...

PingWest品玩 · 今天 04:25

微软Edge浏览器起全面登陆iOS/Android

PingWest品玩 · 昨天 11:03

央视创业帮沉浮录

首席人物观 · 今天 09:00

大家好，这是鹿晗用来介绍女朋友的手机

虎嗅 虎嗅APP · 昨天 22:02 3

新闻 体育 汽车 房产 旅游 教育 时尚 科技 财经 娱乐 更多 登录 狐友

广告 · 今天 11:7

分享到 SOHU.com

三星手机大溃败！专利对抗难敌华为，超62%涉案专利被判无效

李俊慧 · 今天 09:42

科学培育出更逼真的类大脑：直径4毫米微型球状

IT之家 · 今天 08:36

驾驶大数据公司Nauto从微软和谷歌挖来两名高管，加速全球业务扩张

36氪 · 昨天 17:42

谷歌拟用气球向飓风灾区提供紧急网络服务

IT之家 · 今天 09:46

加微信每天推荐3支短线牛股，可以先观察几天看看效果！不花钱

散户必看 抄底 涨停

广告 · 今天 11:7

90岁创业者褚时健：嫌自己太年轻

接招 · 昨天 21:05

109

中国服装企业生死局：打不起的硬仗 输不起的供应链

砍柴网\_SUFFIX\_MP · 今天 08:34

沥青打造新锂电池：可将充电时间缩短十倍以上

IT之家 · 昨天 15:17

2

你可能不知道，自媒体还可以这样赚钱

浩博大数据 · 今天 10:01

3股跌出黄金底，现可大胆抄底！

广告 · 今天 11:7

加载更多