

学界 | 让机器耳濡目染：MIT提出跨模态机器学习模型

2017年06月11日 15:15:12 机器之心

0

选自arXiv

机器之心编译

作者：Yusuf Aytar等人

参与：李泽南

不变性表示（invariant representation）是视觉、听觉和语言模型的核心，它们是数据的抽象结果。人们一直希望在视觉、有噪音的音频、有同义词的自然语言中获取观点和大量不变性表示。具有识别能力的不变性表示可以让机器从大量数据中学习特征，从而获得近似于人类的识别效果。但在机器学习领域，目前这一方面的研究进展有限。

对此，麻省理工学院（MIT）的 Yusuf Aytar 等人最近在一项研究中提出了全新的方法：研究人员通过多种关联信息的输入让机器学习了跨模态数据的通用表达方式。在文字语句「她跳入了泳池」中，同样的概念不仅出现在视觉上，也出现在了听觉上，如泳池的图像和水花飞溅的声音。如果这些跨模态的表示存在关联，那么它们的共同表示就具有鲁棒性。上文中的句子、泳池的图像和水声应当具有相同的内在表示。

论文：See, Hear, and Read: Deep Aligned Representations


链接：<https://arxiv.org/abs/1706.00932>

See, Hear, and Read: Deep Aligned Representations

Yusuf Aytar, Carl Vondrick, Antonio Torralba
Massachusetts Institute of Technology
{yusuf,vondrick,torralba}@csail.mit.edu

摘要

我们利用大量易于获得的同步数据，让机器学习系统学会了三种主要感官（视觉、声音和语言）之间共有的深度描述。通过利用时长超过一年的视频配音和百万条配和图片匹配的句子，我们成功训练了一个深度卷积神经网络对不同信息生成共同的表示。我们的实验证明，这种表示对于一些任务是有效的，如跨模式检索或在形态之间的传递分类。此外，尽管我们的神经网络只经过了图片+文字和图片+声音的配对训练，但它也在文本和声音之间建立了联系——这在训练中未曾接触。我们的模型的可视化效果揭示了大量自动生成，用于识别概念，并独立于模态的隐藏单元。

 机器之心

专业的人工智能媒体与产业服务平台。

热文排行

- 日榜周榜月榜
- 国务院领导说出了大实话：房地产泡沫或..
 - 为什么越是有钱人，越要贷款买房？
 - 30万亿资金正在撤出楼市找韭菜，城市家..
 - 20家银行停止房贷！为拿到贷款，房奴要..
 - 比去年更惨！刚需今年买房要过三道关！
 - 10年前中国房价最高的县城，现在变成
 - 监管层发动组合拳 6月行情风向已变？
 - 旅美台湾人上海归来后说：台湾下一代自..
 - 15股中期业绩暴增 机构新猎物曝光
 - 嘲讽开发商：购房还有刚需？别做梦了韭..



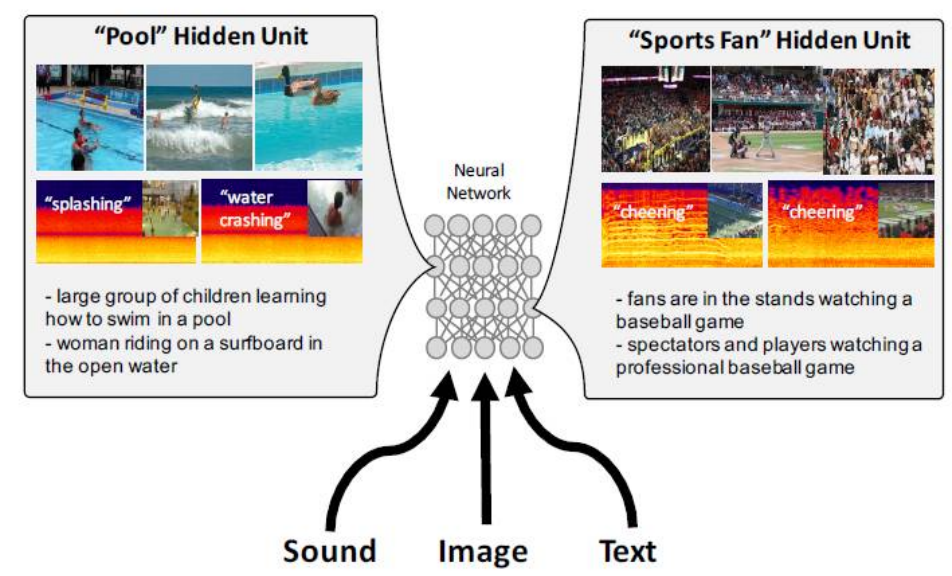


图 1. 共同表示：研究人员提出了深度跨模态卷积神经网络，它可以学习三种表征方式：视觉、听觉和文字阅读。在此之上，研究人员展示了输入信息可以激活网络中的隐藏单元，其中被激发的概念位置独立于模态。

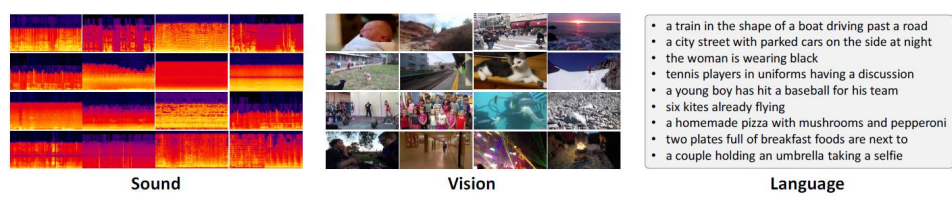


图 2. 数据集：研究人员使用了大量未加工、无约束的数据对概念表达进行训练。

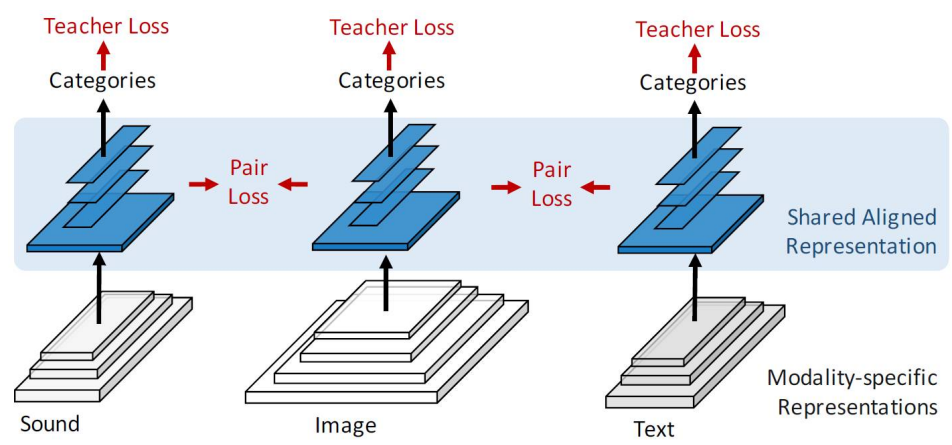


图 3. 学习通用表示方法：研究人员设计了一种能够同时接收图像、声音和文字输入的神经网络。该模型从模态专属表示（灰色）中产生一种通用表示，同时适用于不同模态（蓝色）。研究人员同时使用模型转换损失和配比排名损失来训练这个模型。模态专有层是卷积的，不同模态的共享层则是全连接的。

Input Query	Sound Retrievals		Text Retrievals	Image Retrievals	
			- A dog lying down on the beach - The dog belongs to the homeowner		
			- Steel tracks under the train. - The train platform		
The choppy water the man is riding			- A person stands on water skis in the water - A couple of kayakers paddling through water		

图 4. 跨模式反演示例：MIT 的研究人员展示了使用深度表示，跨声音、图像和文字三种模态的顶层反演

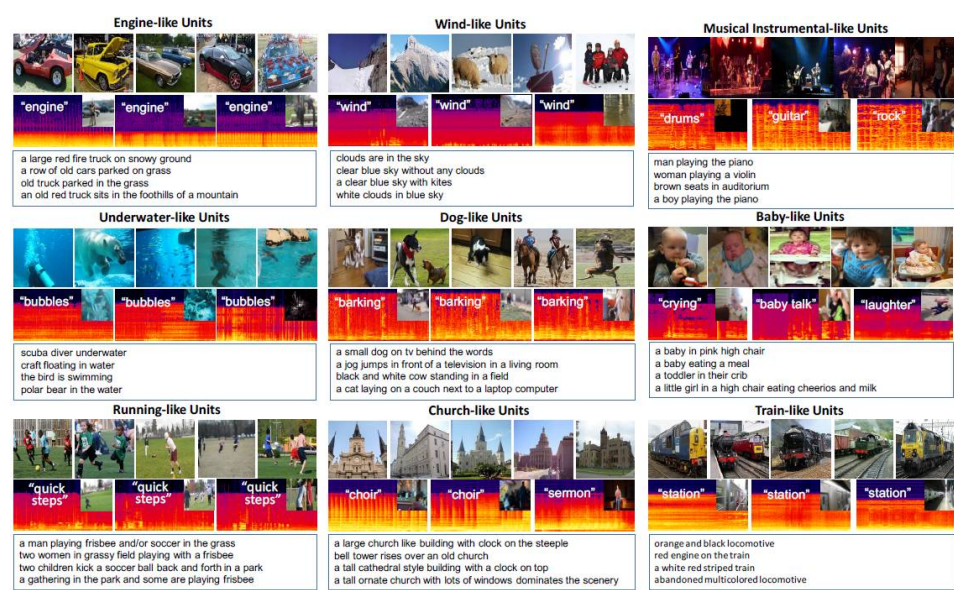


图 5. 隐藏单元的可视化：研究人员通过模型的可视化发现了一些隐藏单元。注意：频谱图（红/黄色的热区显示）之外，还有原始视频和与之对应的描述声音，后者仅用于可视化目的。

不变性表示可以让计算机视觉系统可以在不受约束的、现实世界环境中高效运行。在实验中，研究人员发现了一些联结表达方式具有更高的分类和检索性能，可以应对未遇到过的新情况。麻省理工学院的学者们相信，对于下一代机器感知而言，跨模态的表示具有重要意义。



本文为机器之心编译，转载请联系本公众号获得授权。

加入机器之心（全职记者/实习生）：hr@jiqizhixin.com

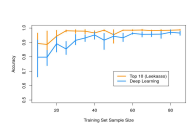
投稿或寻求报道：editor@jiqizhixin.com

广告商务合作：bd@jiqizhixin.com

点击阅读原文，查看机器之心官网↓↓↓

作者历史文章

观点 | 小心训练模型，数据少也可以玩转深度学习



选自Github作者：AndrewL.Beam机器之心编译最近，JeffLeek在SimplyStats 上发表了一篇题为「如果你的数据量不够大就不要使用深度学[详细]

2017年 06月11日 15:15

资源 | PyTorch第一版中文文档发布



机器之心报道参与：黄小天近日，使用GPU和CPU优化的深度学习张量库PyTorch上线了其第一版中文文档，内容涵盖介绍、说明、Package参考、torchvi[详细]

2017年 06月10日 13:15

资源 | 神经网络目标计数概述：通过Faster R-CNN实现当前最佳的目标计



选自SoftwareMill机器之心编译作者：KrzysztofGrajek参与：黄小天在机器学习中，精确地计数给定图像或视频帧中的目标实例是很困难的一个问题。[详细]

2017年 06月09日 11:05

学界 | Facebook 新研究：大批量SGD准确训练ImageNet仅需1小时



选自arXiv机器之心编译参与：蒋思源由于近来互联网数据越来越大，深度学习模型越来越复杂，执行训练的时间也越来越长。因此近日Facebook提出了一种将批量大小提[详细]

2017年 06月09日 11:05

教程 | 如何使用Swift在iOS 11中加入原生物理机器学习视觉模型



选自Hackernoon机器之心编译作者：AlexWulff参与：侯韵楚、李泽南随着WWDC大会上iOS11的发布，苹果终于推出了原生物理学习和机器视觉框架，由[详细]

2017年 06月09日 11:05

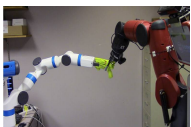
重磅 | 波士顿动力被软银收购，「被丰田收购」传言告破



机器之心报道机器之心编辑部Alphabet（谷歌）想要甩手波士顿动力（BostonDynamics）的传言已经持续了很长时间，而接手者基本上已经被认为是丰田了，[详细]

2017年 06月09日 11:05

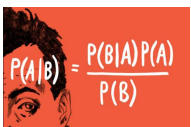
业界 | CMU和谷歌联手研制左右互搏的对抗性机器人



选自IEEESpectrum机器之心编译作者：EvanAckerman参与：蒋思源、SmithCMU和谷歌研究者正在使用基于博弈论和深度学习的对抗性训练策略来提[详细]

2017年 06月09日 11:05

从贝叶斯角度，看深度学习的属性和改进方法



选自arXiv.org机器之心编译参与：蒋思源、吴攀深度学习是一种高效的非线性高维数据处理方法，它可以更自然地解释为一种工程或算法，而本论文希望从贝叶斯的角度将[详细]

2017年 06月08日 12:15

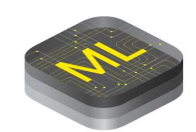
课程 | 来自硅谷的人工智能工程师直通车：打造工业级AI工程师！



人工智能是目前发展最快的领域之一。根据职业发展平台Paysa的专业预测，人工智能在未来5年，将发展成为一个价值160亿美元的市场。仅2017年，美国公司就计划在[详细]

2017年 06月08日 12:15

资源 | 用苹果Core ML实现谷歌移动端神经网络MobileNet



选自GitHub机器之心编译作者：MatthijsHolleman参与：李泽南6月5日开幕的WWDC2017开发者大会上，苹果正式推出了一系列新的面向开发者的[详细]

2017年 06月08日 12:15

- 1

2

3

4

5