

深度学习领域，你心目中 idea 最惊艳的论文是哪篇？

转载

[TomHardy](#) 于 2021-01-31 19:59:51 发布 30 收藏

文章标签：[卷积](#) [人工智能](#) [计算机视觉](#) [深度学习](#) [机器学习](#)

点击上方“计算机视觉工坊”，选择“星标”

干货第一时间送达

来源 | 知乎问答（回答均已授权）

编辑 | 极市平台

导读

深度学习研究中，我们往往会读到让自己觉得想法很惊艳的论文，心中对不同的论文也会有一个排名，那么本文中的提名是你心中的No.1吗？

科研路上我们往往会读到让自己觉得想法很惊艳的论文，心中对不同的论文也会有一个排名，以下介绍了一些知乎作者心中白月光般存在的深度学习领域论文，看看是否你们拥有同样心目中的The one。

提名一

ResNet和Transformer

作者：王晋东不在家

时至今日，许多大领域都离不开这两种结构。**Transformer**更是从**NLP**领域走入了**CV**领域，大有一统天下之势。

ResNet大道至简，更倾向于从原来的CNN结构设计出发，通过大量的实验和分析，添加了skip connection，一招封神。

Transformer则另起炉灶，干脆完全抛弃了RNN的结构，从根本上尝试self-attn加全连接层对于序列建模的能力。

今日的你或许通过实验可以大概搞出来ResNet的skip connection结构，但是能想出来跟transformer一样完全不用RNN、并能让这种当时看来“非主流”的结构work的比RNN还好，就能称得上是天才了。

这其中，固然要有科研的敏锐嗅觉，更多的还是源于超强的代码能力，以及愿意为你这种尝试提供资金和设备支持的大环境。

所以说，要想取得绝对的成功，**天时**（CNN与NLP发展的大环境）、**地利**（所在单位的资源投入）、**人和**（老板与同事的支持），三者缺一不可。

提名二

DUT（视频增稳）

作者：rainy

来分享一篇小众方向（视频增稳/Video Stabilization）的论文，可能不是那种推动领域进步的爆炸性工作，这篇论文我认为是一篇比较不错的把传统方法deep化的工作。

DUT: Learning Video Stabilization by Simply Watching Unstable Videos

<https://arxiv.org/pdf/2011.14574.pdf>

看样子应该是投稿CVPR21，已开源。

<https://github.com/AnnBless/DUTCode>

首先介绍一下视频增稳的定义，如名称所示，视频增稳即为输入一系列连续的，非平稳（抖动较大）的视频帧，输出一系列连续的，平稳的视频帧。

由于方向有点略微小众，因此该领域之前的工作（基于深度学习）可以简单分为基于GAN的直接生成，基于光流的warp，基于插帧（其实也是基于光流的warp）这么几类。这些论文将视频增稳看做了“视频帧生成问题”，但是理想的视频增稳工作应该看做“轨迹平滑”问题更为合适。

而在深度学习之前刘帅成（<http://www.liushuaicheng.org/>）大神做了一系列的视频增稳的工作，其中work的即为meshflow。这里贴一个meshflow解读的链接

<https://www.yuque.com/u452427/ling/qs0inc>

总结一下，meshflow主要的流程为“估计光流-->估计关键点并筛选出关键点的光流-->基于关键点光流得到mesh中每一个格点的motion/轨迹-->进行轨迹平滑并得到平滑后的轨迹/每一个格点的motion-->基于motion得到满足平滑轨迹的视频帧”。

总结了meshflow之后，这篇DUT主要进行的工作其实很简单，在meshflow的框架下，将其中所有的模块都deep化：

LK光流---->PWCNet

SIFT关键点----->RFNet

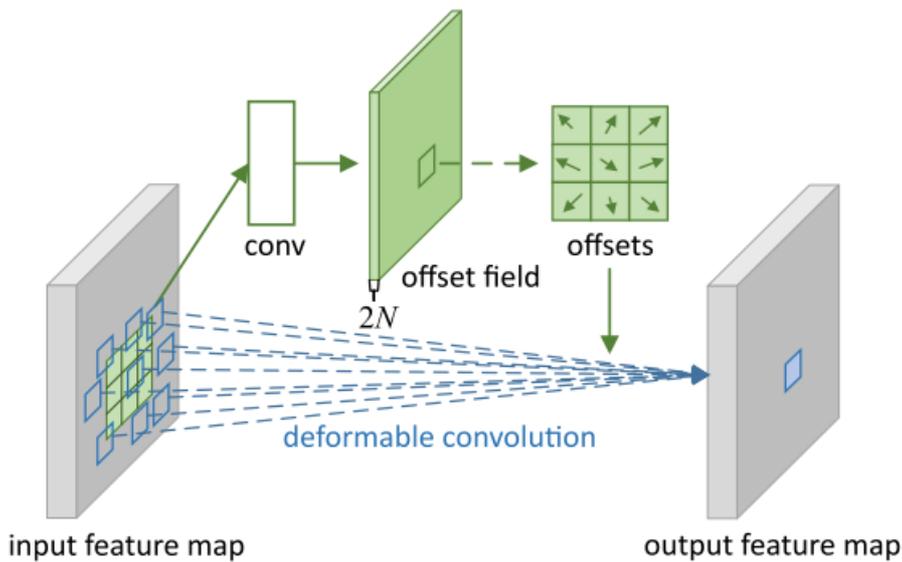
基于Median Filters的轨迹平滑----->可学习的1D卷积

除此之外，由于原始的meshflow是基于优化的方法，因此DUT在替换了模块之后依旧保留了原始的约束项，并且可以使用无监督的方式完成训练，效果也好于一票supervised的方法。

提名三

可形变卷积（DCN）

作者：陀飞轮



当年看Deformable Convolutional Networks(DCN)的时候最为惊艳，可能看过的文章少，这种打破固定尺寸和位置的卷积方式，让我感觉非常惊叹，网络怎么能够在没有直接监督的情况下，学习到不同位置的offset的，然后可视化出来，能够使得offset后的位置能够刚好捕捉到不同尺寸的物体，太精彩了！

提名四

深度学习框架、图像识别、图像生成、模型优化、自然语言处理

作者：叶小飞

我想从深度学习框架、图像识别、图像生成、模型优化、自然语言处理五个领域评选出一篇最惊艳的论文，并且对每一篇论文都赋予一个武侠小说里对应的绝顶武功，以此来表达我的膜拜与狂热。

深度学习框架

论文名称：**Caffe: Convolutional Architecture for Fast Feature Embedding**

论文链接：<https://arxiv.org/abs/1408.5093>

惊艳理由：在那个大家都用matlab和自己diy深度学习框架的年代，贾大神的Caffe横空出世，为深度学习领域创立了一个通用、易拓展的框架，使复现、开发各种新型算法变得更加容易，可以说是开山鼻祖。

对标武功：《天龙八部》内的易筋经。易筋锻骨，重塑七经，这不正和caffe的效用不谋而合？caffe不就相当于重塑了深度学习的筋骨，使得后续各种五花八门的算法变得可能实现？

图像识别

论文名称：Deep Residual Learning for Image Recognition

论文链接：<https://arxiv.org/pdf/1512.03385.pdf>

惊艳理由：一个简简单单的skip connection一招制敌，优雅至极，直接将CNN的表现提升了一个大档次。

对标武功：《天涯明月刀》里傅红雪的刀法。傅红雪的武功就一招——拔刀收刀，简单却致命，正如resnet的skip connection. 这一刀是傅红雪拔了千万次刀后凝练的精魂，正与skip connection是作者做了无数实验与分析后凝练的结构如出一辙。

图像生成

论文名称：Generative Adversarial Networks

论文链接: <https://arxiv.org/abs/1406.2661>

惊艳理由: Encoder-decoder 出现已久, 分类器出现也很久, Goodfellow却是真正意义上把这二者完美结合起来的第一人, 是现如今图像视频模拟生成的鼻祖。

对标武功: 《射雕英雄传》里的左右互搏。老顽童让左手和右手打架, 结果两只手突飞猛进, Goodfellow让generator和discriminator互相打架, 结果两个模型变得越来越强, 最后甚至可以以假乱真。

模型优化

论文名称: Distilling the Knowledge in a Neural Network

论文链接: <https://arxiv.org/pdf/1503.02531.pdf>

惊艳理由: 知识蒸馏的开山之作, 在不增加任何online inference资源的情况下, 让模型得到极大优化。

对标武功: 《天龙八部》里的北冥神功。段誉吸各个高手的内功变成了天龙三绝之一, student net吸取teacher net的知识变成了更robust的模型。

自然语言处理

论文名称: Language Models are Few-Shot Learners (GPT-3)

论文链接: <https://arxiv.org/pdf/2005.14165.pdf>

惊艳理由: 在看到这篇论文之前, 我做梦也想不到一个NLP模型居然有175亿个参数, 可以说是深度学习里的暴力美学的极致了。

对标武功: 降龙十八掌。降龙十八掌刚猛无双, 遇到强敌以刚劲的掌力与无所畏惧的气势压倒对方。GPT-3庞大无比, 遇到语言数据以175亿的模型参数与超出想象的计算资源死磕硬刚。如果乔峰是个深度学习科学家, 一定会爱死这个模型。

提名五

CAM, class activation map

作者: Feras

那就从我的研究领域中挑一个出来吧, 我的研究方向是基于image-level的弱监督语义分割, (貌似这个点近两年趋势渐淡), 而其中令我最惊艳的就是CAM, class activation map

文章题目叫Learning Deep Features for Discriminative Localization, google百度一下都可以找到。这篇文章其实是想探究我们的CNN在学习图像的时候到底重点关注在哪个部分。这里抛开论文里面的繁琐的数学解释啥的(大家可以看看原论文), 最后论文用一张图表示了这个问题大概是一个过程。

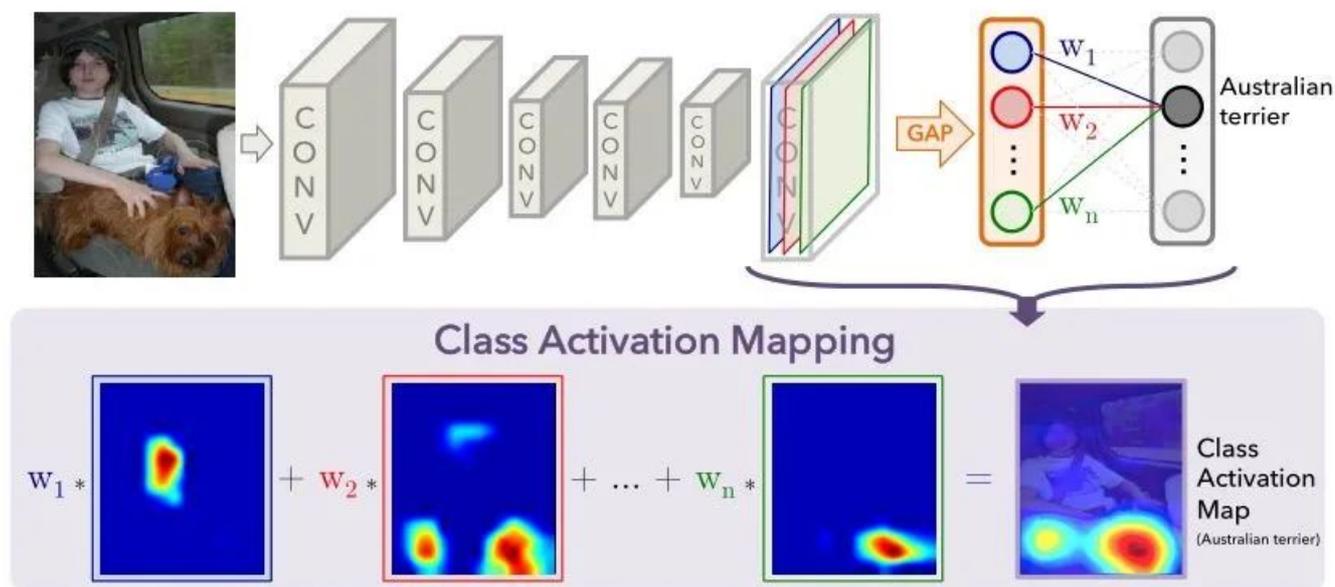


Figure 2. Class Activation Mapping: the predicted class score is mapped back to the previous convolutional layer to generate the class activation maps (CAMs). The CAM highlights the class-specific discriminative regions.

对你没有看错，图像关注的部分就是将该类的fc层中的权重和feature maps对应加权求和就行了。说实话我觉得这个真的是经过很多实验才发现的idea。因此通过这个CAM我们便可知这个网络到底在学什么东西。

至于后面CAM变体例如grad-cam等大家可以去查阅了解。通过这个惊艳的CAM，我觉得是开了基于弱监督图像分割领域的先河，简直是祖先级别的神工作。

为什么这么说呢，基于image-level的弱监督分割旨在仅通过分类标签而生成对应的分割标签图，（毕竟手工标记分割图上的像素太烧钱了呀哈哈）你看看CAM，如果通过阈值一下的话，那些热点处的不就可以作为置信度高的前景像素标签了嘛！！

于是你便可以看到大量的弱监督领域分割之作都是在这个CAM之上完成的。不仅如此，CAM也在可解释领域中被作为一种基本的工具。这篇五年前的文章至今仍在视觉领域中放光发热，让很多的学者以此为基石展开研究。

我也是很感谢这篇工作让我接触到弱监督领域。毕竟是我转做计算机视觉读的第一篇文章hhhh，所以，thank you, CAM!

参考链接：

回答1-王晋东不在家：

<https://www.zhihu.com/question/440729199/answer/1697212235>

回答2-rainy：

<https://www.zhihu.com/question/440729199/answer/1693346011>

回答3-陀飞轮：

<https://www.zhihu.com/question/440729199/answer/1695810150>

回答4-叶小飞：

<https://www.zhihu.com/question/440729199/answer/1698687630>

回答5-Ferenas：

<https://www.zhihu.com/question/440729199/answer/1695809572>

你心目中 idea 最惊艳的论文是哪篇？欢迎在下方留言~

本文仅做学术分享，如有侵权，请联系删文。

下载1

在「计算机视觉工坊」公众号后台回复：**深度学习**，即可下载深度学习算法、3D深度学习、深度学习框架、目标检测、GAN等相关内容近30本pdf书籍。

下载2

在「计算机视觉工坊」公众号后台回复：**计算机视觉**，即可下载计算机视觉相关17本pdf书籍，包含计算机视觉算法、Python视觉实战、Opencv3.0学习等。

下载3

在「计算机视觉工坊」公众号后台回复：**SLAM**，即可下载独家SLAM相关视频课程，包含视觉SLAM、激光SLAM精品课程。

重磅！计算机视觉工坊-学习交流群已成立

扫码添加小助手微信，可申请加入3D视觉工坊-学术论文写作与投稿 微信交流群，旨在交流顶会、顶刊、SCI、EI等写作与投稿事宜。

同时也可申请加入我们的细分方向交流群，目前主要有**ORB-SLAM系列源码学习**、**3D视觉**、**CV&深度学习**、**SLAM**、**三维重建**、**点云后处理**、**自动驾驶**、**CV入门**、**三维测量**、**VR/AR**、**3D人脸识别**、**医疗影像**、**缺陷检测**、**行人重识别**、**目标跟踪**、**视觉产品落地**、**视觉竞赛**、**车牌识别**、**硬件选型**、**深度估计**、**学术交流**、**求职交流**等微信群，请扫描下面微信号加群，备注：“研究方向+学校/公司+昵称”，例如：“3D视觉+上海交大+静静”。请按照格式备注，否则不予通过。添加成功后会根据研究方向邀请进去相关微信群。原创投稿也请联系。



▲长按加微信群或投稿



▲长按关注公众号

3D视觉从入门到精通知识星球：针对3D视觉领域的知识点汇总、入门进阶学习路线、最新paper分享、疑问解答四个方面进行深耕，更有各类大厂的算法工程人员进行技术指导。与此同时，星球将联合知名企业发布3D视觉相关算法开发岗位以及项目对接信息，打造成集技术与就业为一体的铁杆粉丝聚集区，近2000星球成员为创造更好的AI世界共同进步，知识星球入口：

学习3D视觉核心技术，扫描查看介绍，3天内无条件退款

3D与SLAM

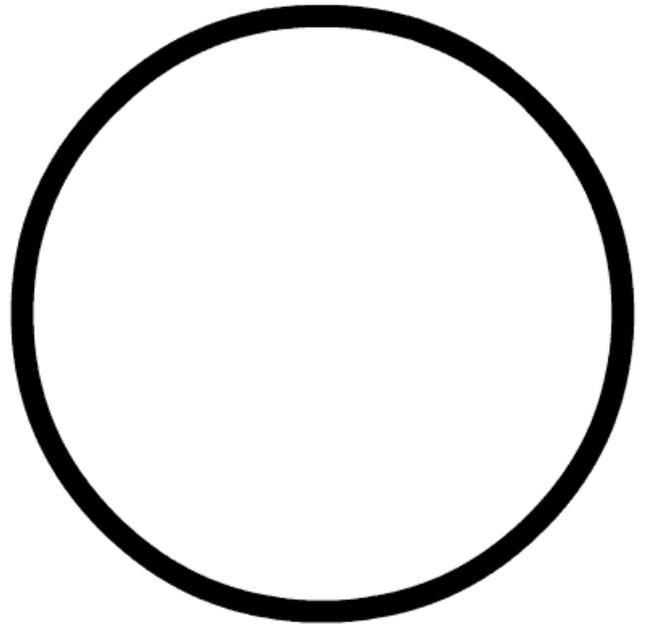
3D视觉从入门到精通

星主：小凡

知识星球
微信扫码预览星球详情

A green rectangular promotional card for a knowledge planet. On the left, there is a circular logo with the text '3D与SLAM'. To the right of the logo, the title '3D视觉从入门到精通' is written in large white characters, with '星主：小凡' below it. At the bottom left, there is the '知识星球' logo and the text '微信扫码预览星球详情'. On the right side of the card, there is a QR code.

圈里有高质量教程资料、可答疑解惑、助你高效解决问题



觉得有用，麻烦给个赞和在看~