

[首页](#)[专栏](#)[专题](#)[公开课](#)[AI慕课学院](#)[爱搞机](#)[极客购](#)[申请专栏作者](#)[业界](#)[GAIR](#)[人工智能](#)[智能驾驶](#)[网络安全](#)[AR/VR](#)[机器人](#)[Fintech](#)[物联网](#)[未来医疗](#)[智能硬件](#)[AI+](#)[开发者](#)[人工智能](#)[正文](#)

## 大牛讲堂 | 深度学习Sequence Learning技术分享

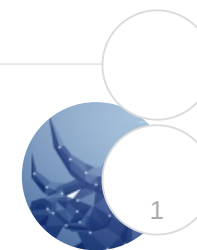
本文作者：大牛讲堂

2016-08-22 19:35

导语：本文主要介绍了深度学习中Sequence Learning技术的分享，也介绍了其对传统光学框架的改造。

雷锋网按：本文作者都大龙，2011年7月毕业于中科院计算技术研究所；曾任百度深度学习研究院（IDL）资深研发工程师，并连续两次获得百度最高奖——百万美金大奖；现在Horizon Robotics负责自主服务机器人、智能家居以及玩具方向的算法研究与开发，涉及深度学习、计算机视觉、人机交互、SLAM、机器人规划控制等多个领域。

### 深度学习独领风骚



大牛讲堂

知情人士

嵌入式人工智能全球领导者

人工智能领域深度学习独领风骚自2006年Geoffery Hinton等在《科学》(Science)杂志发表那篇著名的论文开始，深度学习的热潮从学术界席卷到了工业界。

从那天起，深度学习在工业界的应用就如火如荼，真正开始“深度”影响我们的生活。比如这个大牛同学，参与开发了国内最早的基于CDNN的图像识别技术，极大提升了计算机视觉相关线上应用的效果，还创新性的主导研发了基于CNN和BLSTM的OCR识别系统，大幅提升了商用OCR系统的识别率，他的工作影响了包括你我在内的千万互联网用户。

## 深度学习的优势在哪里

人工智能的特征之一是学习的能力，即系统的性能是否会随着经验数据的积累而不断提升。我们认识到深度学习主要在以下三个方面具有巨大优势：

- 1.从统计和计算的角度看，深度学习特别适合处理大数据。在很多问题上，深度学习是目前我们能找到的最好方法。
- 2.深度学习不是一个黑箱系统。它提供一套丰富的、基于联接主义的建模语言（建模框架）。利用这套语言系统，我们可以表达数据内在的丰富关系和结构，比如用卷积处理图像中的二维空间结构，用递归神经网络(Recurrent Neural Network, RNN)处理自然语言等数据中的时序结构。
- 3.深度学习几乎是唯一的端到端机器学习系统。它直接作用于原始数据，自动逐层进行特征学习，整个过程直接优化某个目标函数。

## 关于Sequence Learning的技术分享现场实录

关于Sequence Learning的技术分享现场实录从2012年的ImageNet竞赛开始，深度学习首先在图像识别领域发挥出巨大威力。随着研究的深入，深度学习逐渐被应用到音频、视频以及自然语言理解领域。这些领域的特点是针对时序数据的建模，我们将其称之为Sequence Learning。如何利用深度学习来进行端到端的学

[发私信](#)[当月热门文章](#)

### 最新文章

Facebook要用AI提前探测恐怖主义，到哪一步了？

你真的需要机器学习吗？

谷歌推出基于注意机制的全新翻译框架，Attention is All You Need!

哈工大在CoNLL上斩获全球第四，车万翔博士详解背后的技术细节

监督学习×强化学习，Facebook让聊天机器人学会谈判

重磅 | DeepMind谷歌学会关系推理，还击败了AlphaGo

### 热门搜索

[小米](#)[iPhone5](#)[移动应用](#)[社交网络](#)[云服务](#)[ARM](#)[陌陌](#)

习，并摒弃基于人工规则的中间步骤，以提升Sequence Learning的效果已经成为当前研究的热点。

开源硬件

黄仁勋

Macbook Air

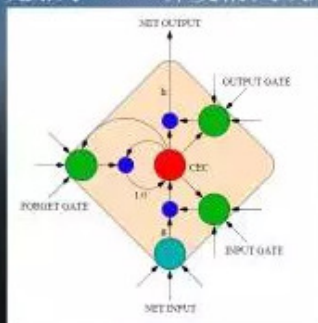
Cortana

Sequence Learning 已经成功应用到多个领域，比如语音识别、Image Caption、机器翻译、OCR 等，它们的共同特点是利用 DNN 或者 CNN 提取高级语义特征，利用 RNN 建模时序信息。在损失函数方面，除了常见的 logistic 损失外，还引入了结构化损失，比如 CTC 等序列对序列的损失等。

## LSTM (Long Short temporal Memory)

- 特点

- 具有特定结构的Recurrent类型
- 通过一系列Gate的开关来自适应的建模长时的信息
- 克服了SRNN梯度消失或爆炸的问题



simple RNN的变体—LSTM

## CTC (Connectionist Temporal Classification)

- CTC removes the need of pre-segmented data, and estimates the probabilities of the complete label sequences overall
- Post-probability of the path  $\pi$  on the given observation  $\mathbf{x}$  and ground-truth label  $S$

$$p(\pi|\mathbf{x}, S) = \prod_{t=1}^T y_{\pi_t}^t$$

- Blank and repeated pattern removal

many-to-one map  $\mathcal{B} : L^{tT} \mapsto L^{\leq T}$

### CTC结构化损失函数

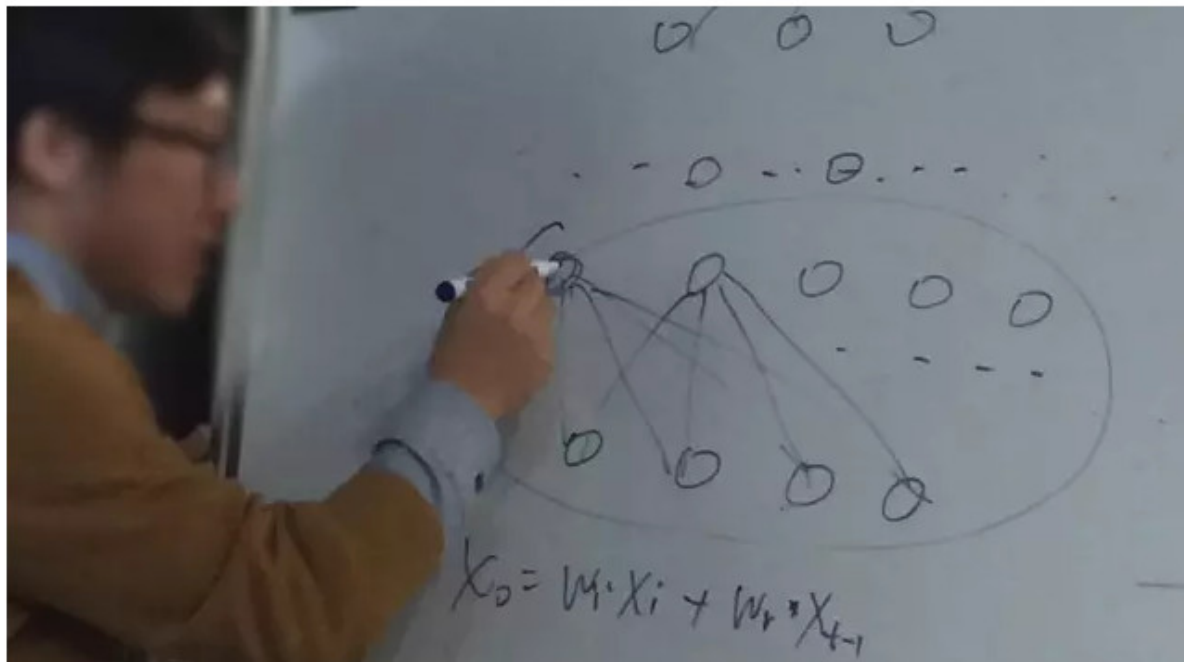
在 Sequence Learning 中，我们认为 RNN 和 sequence 相关的结构化损失函数是当前时序学习取得巨大成功的重要组成部分。除了传统的 simple RNN 外，出现了很多 RNN 变种，比如LSTM（Long Short Temporal Memory）、GRU（Gated Recurrent Unit）等，已经被广泛的应用到时序学习的任务中，它们都具有特定的 Recurrent 结构，并且通过一系列 gate 开关自适应的建模长时的信息，都在一定程度上克服了 Simple RNN 优化过程中的梯度消失或爆炸问题。CTC 作为一种结构化损失函数，它没有必要对序列数据进行分割，并且估计整体序列标注概率作为损失，已经被广泛应用到 OCR、语音识别以及其他序列识别任务中。

在这里他就以 OCR 为例，介绍如何利用机器学习尤其是 Sequence Learning 技术对传统 OCR 技术进行改造升级。



# 对传统光学字符识别框架的改造

## 基于端到端的序列学习



大牛正在白板上讲解RNN

光学字符识别的概念早在20世纪20年代便被提出，一直是模式识别领域研究中极具代表性的重要课题。

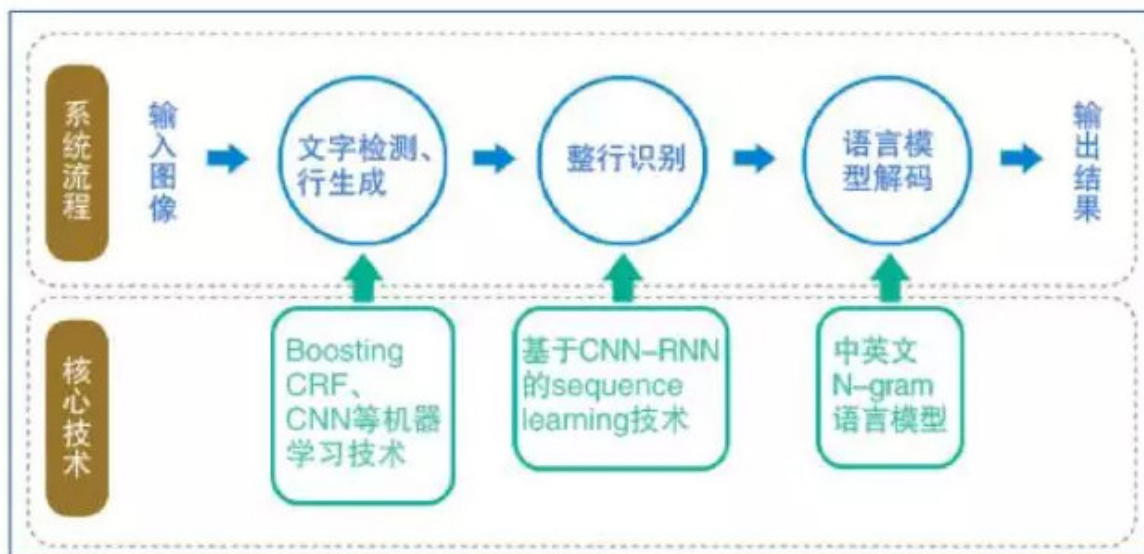
经典的光学字符识别系统从输入图像到输出最终的文字识别结果，历经版面分析、行分割、字分割、单字识别、语言模型解码和后处理。涉及的技术分为基于经验制定的规则和基于统计学习的模型两大类。前者包括系统预处理阶段（版面分析、行分割、字分割）的二值化、连通域分析、投影分析等，以及后处理阶段的规则噪声过滤器；后者包括基于类方向梯度直方图（Histogram

of Oriented Gradient, HOG) 特征的单字识别引擎和基于N-gram 的语言模型, 用于单字识别和语言模型解码阶段。

在数据简单、条件可控的情况下, 经典的光学字符识别技术架构通过细致的人工规则制定和适量的模型参数学习, 便可以达到比较理想的识别精度。但在广泛的自然场景中, 文字呈现出的图像信息复杂度显著增大, 而拍摄图像的条件又得不到很好的控制, 经典的光学字符识别技术架构难以满足实际应用的需求。究其原因, 是这一技术架构的处理流程繁琐冗长导致错误不断传递, 以及过分倚重人工规则并轻视大规模数据训练。

## 解决方案

针对复杂场景的特点和经典技术框架的不足, 利用机器学习尤其是 Sequence Learning 技术对光学字符识别的系统流程和技术框架进行了大幅改造。





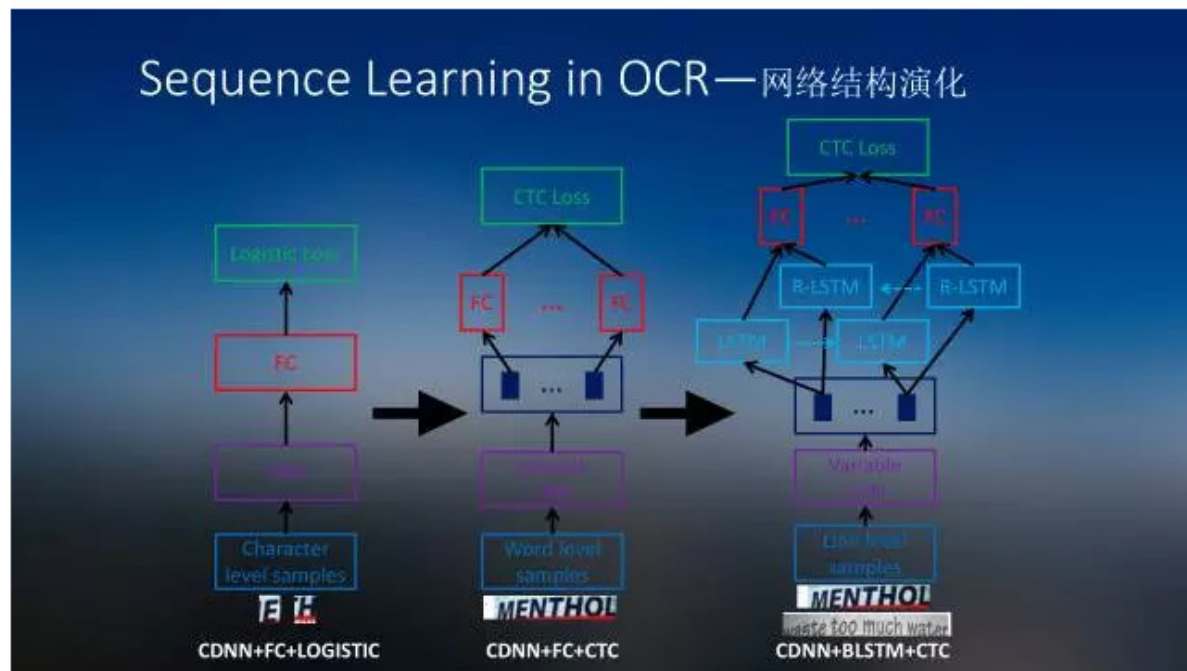
在系统流程方面，摒弃传统的二值化和连通域等基于规则的方法，引入基于学习的 Boosting 文字检测概念，并和行分割合并成新的预处理模块，任务是检测图像中包含文字的区域并生成相应文字行；将字分割和单字识别合并成新的整行识别模块；基于 N-gram 的语言模型解码模块予以保留，但将主要依赖人工规则的版面分析和后处理模块从系统中删除。6 个步骤减少到3 个步骤，降低了误差传递造成的不良影响。

此外，由于整行文字识别是一个序列学习（Sequence Learning）问题，我们有针对性地研发出基于双向长短期记忆神经网络（Bidirectional Long Short-term Memory, BLSTM）的递归神经网络序列模型学习算法，结合卷积神经网络模型提取出的图像特征，不考虑每个字符出现的具体位置，只关注整个图像序列对应的文字内容，使得单字分割和单字识别问题融为一体，最终实现深度学习理论追求的理想——端到端训练。

## Sequence Learning in OCR—设计思路

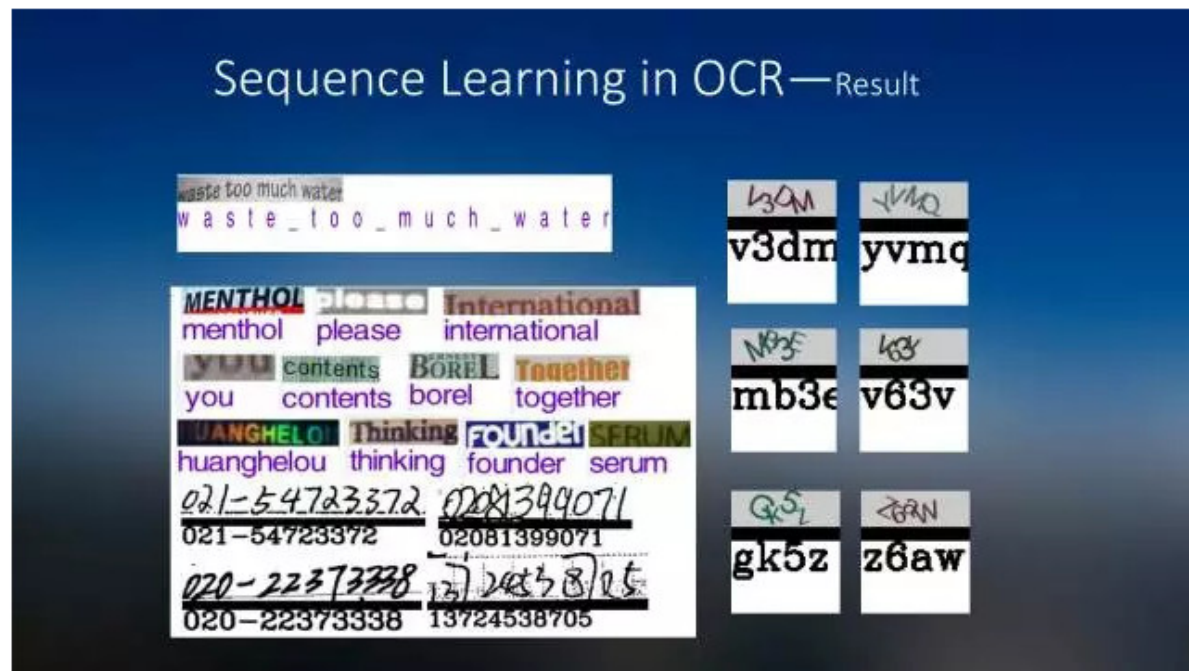
- CNN
  - 强大的图像特征提取器
  - 提取高级的语义信息
  - 对图像扰动具有一定的容忍性
- RNN-BLSTM (Bidirectional Long Short Temporal Memory)
  - 强大的序列学习器
  - 灵活的建模序列内部关系
  - 对序列的局部失真具有一定的鲁棒性
  - LSTM
    - 具有特定结构的Recurrent类型
    - 通过一系列Gate的开关来自适应的建模长时的信息
    - 克服了SRNN梯度消失或爆炸的问题
- Sequence-to-Sequence Loss
  - CTC Loss (Connectionist Temporal Classification)
  - 直接计算序列输出与序列ground truth的损失





这样做能够充分利用文字序列上下文关联进行消歧，避免传统方法中字符分割造成的不可逆转的错误。这一序列学习模型极其擅长识别字分割比较困难的文字序列，甚至包括潦草的手写电话号码。此外，这一序列学习模型还使得训练数据的标注难度大为降低，便于收集更大规模的训练数据。不同语言（即便字词、句子的长短结构迥异）光学字符识别问题也可以纳入同一个技术框架内统一解决，大幅降低系统维护成本。





## 总结与展望

作为深度学习以及 Sequence Learning 的实践者，我们获得许多颇具价值的经验和知识：

1. 丰富的图像扰动是我们将关于图像的先验知识用于深度学习输入端的有效手段。不同于许多其他数据，图像和视频在时间、空间维度上具有良好的连续性和结构性，且包含大量冗余信息。无论使用平移和翻转，还是旋转、缩放、高斯和椒盐噪音、错切等图像处理变换，都能够产生大量有效的训练数据，增强深度学习模型的鲁棒性。
2. RNN作为一种序列信息的建模语言能够有效的建模序列内部的依赖关系。RNN可以利用它内部的记忆来处理任意时序的输入序列，极大的降低了视频处理、语音识别、语义理解中序列建模



的难度。

3.结构化损失函数是我们将模型化知识用于深度学习输出端的有效方式。采用人工模型对深度学习模型输出进行后处理时，具有针对性的结构化损失函数往往能够帮助深度学习过程更快地收敛到更加理想的状态。

展望未来，基于深度学习的序列识别问题，可围绕如下重点展开：

## 增强学习

与卷积神经网络和递归神经网络相比，增强学习产出的模型能够根据数据特点更灵活地产生输入序列，并通过更加模糊的监督方式进行模型训练。这样可以精简模型的复杂度，提高预测速度，同时大幅降低训练数据的标注难度，使得学习和预测过程不需要过多的人工参与，形式上更接近真正智能的学习模式。

## Attention modelAttention

作为一种抽象的概念，它模拟人的识别行为，并不只是利用序列当前时刻的状态信息，而是在decode的过程中对之前的序列状态信息进行自适应的建模加权，从而得以利用context的全部信息。

雷锋网按：本文由大牛讲堂授权雷锋网(公众号：雷锋网)发布，如需转载请联系原作者。

雷锋网原创文章，未经授权禁止转载。详情见[转载须知](#)。





3人收藏

分享：



## 相关文章

AI科技评论

深度学习

序列学习

中科院

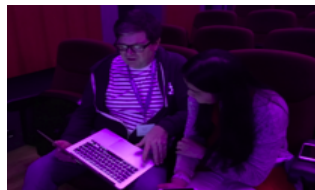
ImageNet



AI科技评论招人啦，学术主编、学术编辑、外翻编辑虚



实例揭秘：ICRA上的学术界和工业界是怎样殊途同归



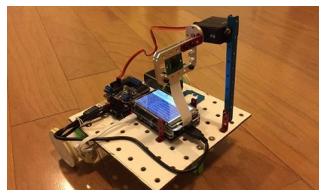
AI科技评论专访Yann LeCun: 关于深度学习未来



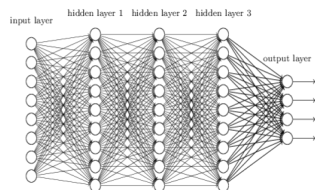
厉害了！数据科学杯：10000名数据科学家参加，



腾讯布局保险又下一城，BAT为何都看好保险领域？



Google工程师：教你用树莓派+Arduino+TensorFlow



从这开始了解深度学习——视觉的深度学习与网络



被微软收购后，Maluuba着手研究多重引导强化学习机



文章点评：

我有话要说.....

☐ 同步到新浪微博

提交

最新评论

 李宗仁 08月22日 23:44

从统计和计算的角度看，深度学习特别适合处理大数据。—————一语成X

(0) 回复

热门关键字

热门标签 微信小程序平台 微信小程序在哪 CES 2017 CES 2016年最值得购买的智能硬件 2016 互联网 小程序 微信朋友圈 抢票软件 智能手机 智能家居 智能手环 智能机器人 智能电视 360智能硬件 智能摄像机 智能硬件产品 智能硬件发展 智能硬件创业 黑客 白帽子 大数据 云计算 新能源汽车 无人驾驶 无人机 大疆 小米无人机 特斯拉 VR游戏 VR电影 VR视频 VR眼镜 VR购物 AR 直播 扫地机器人 医疗机器人 工业机器人 类人机器人 聊天机器人 微信机器人 微信小程序 移动支付 支付宝 P2P 区块链 比特币 风控 高盛 人脸识别 指纹识别 黑科技 谷歌地图 谷歌 IBM 微软 乐视 百度 三星s8 腾讯 三星Note8 小米MIX 小米Note 华为 小米 阿里巴巴 苹果 MacBook Pro iPhone Facebook GAIR IROS 双创周 云栖大会 智能硬件公司 智能硬件 QQ红包 支付宝红包 敬业福 支付宝敬业福 支付宝集五福 Waymo 虚拟现实 深度学习 人工智能 中国银联 蚂蚁金服 WRC CNCC 云安全 upright openwrt eero 人工智能 深度学习 能打电话的手表 ipad kindle 撸一 做产品 蓝牙遥控器方案 hyperloop one 美国大选 实时统计 小米mix 设计师 理财推荐 令计划 乐视 更多

---

[联系我们](#)   [关于我们](#)   [加入我们](#)   [意见反馈](#)   [投稿](#)

---

Copyright © 2011-2018 www.leiphone.com 雷锋网-移动互联网智能终端第一媒体 All Rights Reserved 粤ICP备11095991号-1



ICP证粤B2-20150332

