

# CCF – CV Series Lectures

## 计算机视觉前沿技术及应用

今天上午参加了 CCF-CV 走进高校系列报告会第 19 期，

第一个报告学者是中科院自动所的王亮，题目为《深度学习及视觉模式分析中的应用》，重点介绍了讲者应用深度学习技术在视觉数据分析理解方面的工作。自己感兴趣的有以下几个。

接下来是他们做的一些工作：

广义自编码器：为了保持学习到数据表示的局部结构关系，在重构过程中考虑数据之间的相似性关系，基于这样的思路，将传统的 Autoencoder 改进为广义的 Autoencoder。

深度关系学习：提出一种新的监督学习算法：High-order Boltzmann Machines。这个可以应用在关系学习的任务中，如人脸匹配。从传统的 RBM 到 Gated CHBM，再在 Gated CHBM 之间加入四个分解量：Factored Gated RBM。

讲者分享了 RNN 最近在计算视觉领域的一些应用，因为最近在用 RNN，所以看上去特别亲切。在视觉领域，RNN 从 2014 年开始被大量应用，一个是结合 CNN 做 image 和 video 的 caption，另一个是扩展到多维如用 2 维的 LSTM 做场景标注，3 维 LSTM 做大脑扫描图像分割等等。

一个是用 BiRNN 分析人体骨骼行为，将人体骨架分解为一个一个的 part，利用局部轨迹建模，然后融合，将融合的结果再进行建模，再做融合，再建模，最后融合成一个人体骨骼运动模型，进行动作识别。至于模型为什么是这样的，有很大一部分是通过实验做出来的结果。

然后是跨视角的步态识别，这个运用了网络的约束：一是融合权重特征的约束，利用不同维度做约束，利用其中的关联关系，一个是语义上的约束，在 label 上做关联。

下面一个工作是多帧超分辨率，主要是以下两个方面：

Model long-term temporal dependency of the video sequence.

Replace all full connections with weight-sharing convolution.

将前馈，卷积和循环结合在了一起使用。

再一个比较有趣的是反馈卷积神经网络，减少网络的连接，只保留与图像特定目标相关的连接。

关于未来的一些方向：基于类脑智能的研究，当前计算机视觉中的很多机制都是基于最前沿的生物学机制研究成果。

从王亮研究员的一些分享可以看出，他们组将各种 RBM、Autoencoder、RNN、LSTM 等用在了解决视觉问题中。他们组运用了大量的关系、分解融合等思路对传统的模型进行了改进。能做出这些工作，要对这些基本模型有相当深入的理解，并且能够实现各种模型和想法。

第二位讲者是复旦大学的姜育刚教授，他分享的题目是 Video Content Recognition with DeepLearning。主要包含这两方面的内容：一个是数据集一个是模型，重点讲解了他们自己的数据集的构建，和自己参与组织的比赛，目前来说，数据集就是做计算机视觉研究的基础，更好的了解数据集的特点可以更好的展开工作。

Video Classification with Regularized DNN。在深度神经网络中间，加入了两个 regularization，这么做是为了发现内部特征和内部类别之间的关系。

Hybrid Deep Learning Framework：这个是将视频的时空线索分别输入到不同的网络模型中，分别获得固定视频帧的特征及时序特征，局部运动特征及其时序特征，将最后的这些特征做一个融合可以得到一个不错的效果。

Object-Scene Semantic Fusion Network：这个主要是将视频帧中的目标、场景、通用特征等做一个融合，考虑了这些因素进去可以更好的进行建模解决这个问题。

结合最近在做的事情主要有以下收获：

场景分类任务与姜教授介绍的那个工作很类似。这个也跟郑老师进行了讨论，理解了场景才能更好的对场景进行分类，所以一个思路是检测出场景中有什么，然后同时考虑目标和背景再进行分类，这个之前也想到过，但是执行力弱的惊人，一直没有去做。这是在平时的工作中，最迫切需要改进的一点，这也是做事情最需要的特点，能想到的人和能做到的人之间的差距简直是天壤之别。

从王亮研究员做的很多工作中也可以看得出来，他们的成果是基于大量的实验研究的，在掌握神经网络的基本模型之后，在原有的基础上进行实验，针对不同问题进行改进，进行模型的融合等，思路其实并不是很难，肯定会有很多人想到，但重点是他们将想法实现了，这才是最关键的。

这两位讲者对自己接下来怎么使用模型解决问题提供了一个不错的思路

还有一点，最近在研究 LSTM，关于之前看的 Fined-grained classification 问题，有这样一个想法，可以将 2D LSTM 用在这个问题上，LSTM 更能将图像的细节特征考虑进去，这是因为 LSTM 是逐像素进行计算来提取特征的。