

# CV 课程提纲

Jinna Cui

2017.4.20

## Chapter 1

---

### 分类

---

#### 1.1 图像分类的定义

#### 1.2 数据集

图像分类常用的几个数据集: MNIST, CIFAR-10, Caltech-256, PASCAL VOC, ImageNet-1000, MS COCO, Google Open Images dataset.

#### 1.3 分类问题的评价指标

overall accuracy, Kappa 系数, 混淆矩阵, top-1 error, top-5 error

#### 1.4 图像分类的过程

预处理 + 特征提取 + 特征表达 + 特征选择/降维 + 分类器

#### 1.5 图像分类的方法

##### 1.5.1 传统机器学习方法

预处理: 图像增强, 图像复原, 图像分割等

特征提取: **SIFT**, SURF, **HOG**, LBP, FAST, LoG, DoG,

特征表达: **Fisher 向量**, **向量量化**, 软量化, FMM, LCC

特征选择: 按搜索测量进行特征选择的方法, 按评价准则进行特征选择的方法

降维: PCA, LDA, LLE, 特征哈希

分类器: SVM, **K 近邻**, 随机森林, adaboost

### 1.5.2 传统机器学习方法的优点及存在的问题

### 1.5.3 深度学习方法

#### 1.5.3.1 CNN 的优点

介绍卷积神经网络在分类问题上取得的成就，并对比模式识别课及人工神经网络课程学过的 MLP 来从卷积神经网络的结构上解释为何卷积神经网络更适合解决分类问题。

Experiments: LeNet 在 MNIST 数据集上训练，主要为引入卷积神经网络分类

#### 1.5.3.2 CNN 结构

进一步解释 dropout 层和 softmax 层对于分类问题的重要性（李娜会讲卷积层，pooling 层的运算）以及卷积神经网络是如何实现从特征提取到分类的过程。

Experiments: AlexNet 在 MNIST 数据集和 CIFAR-10 数据集上的训练，重点观察 AlexNet 在分类问题上的提升，并对比有无 dropout 层来看 dropout 层对分类问题的重要性。

#### 1.5.3.3 典型 CNN 网络

结合几个典型的卷积神经网络（VGG, GoogLeNet, ResNet）来介绍卷积神经网络的发展趋势。

Experiments: MNIST 数据集和 CIFAR-10 数据集上训练 AlexNet, VGG, ResNet 的实验，并结合五个评价指标对比其结果

### 1.5.4 深度学习方法的优点及存在的问题

## Chapter 2

---

## 细粒度分类

---

### 2.1 细粒度图像分类定义

### 2.2 数据集

细粒度分类常用的数据集：CUB-Bird, StandFord Dog, Stanford Car, OxfordFlower, UEC-FOOD

### 2.3 图像分类方法应用于细粒度分类问题

图像分类方法在细粒度分类问题上表现一般

### 2.3.1 传统机器学习方法在细粒度问题上效果很差

SIFT+BoW+SVM 准确率约 10%。

### 2.3.2 卷积神经网络在细粒度分类问题上的效果不够好

AlexNet, VGG 等准确率低于 70%。

Experiments: AlexNet 和 VGG 在 CUB-Bird 和 Stanford Dog 数据集上的表现。

### 2.3.3 效果不好的原因

(1) 细粒度分类图像中具有区分度的信息往往存在于细小的局部；(2) 细粒度分类的数据集样本数量太少。

### 2.3.4 fine-tune 及数据增强

介绍 fine-tune 及数据增强在细粒度分类问题中的重要性。

Experiments: AlexNet 和 VGG 在 CUB-Bird 和 Stanford Dog 数据集上经过 ImageNet 模型 fine-tune 后的表现。

## 2.4 细粒度图像分类的过程

获得具有有效信息的局部区域 + 特征提取 + 特征表达 + 分类器

## 2.5 细粒度图像分类的方法

### 2.5.1 常用的人工标注信息

boundingbox, part annotation, segmentation 等

### 2.5.2 强监督算法

#### 2.5.2.1 part R-CNN

part R-CNN 除了用 R-CNN 做检测之外，还用到了 AlexNet 做特征提取以及 SVM 做为分类器，方法比较典型且分类准确率相对较高：在测试集没有标注信息的情况下，获得 73.9% 的准确率。

#### 2.5.2.2 pose normalization

用到了姿态对齐以及提取不同层级特征的方法，且分类准确率最高，在 CUB-Bird 数据集上，在训练集和测试集都用标注信息的情况下获得 85.4% 的准确率（最高）。测试集没有标注信息时的准确率为 75.7%。

### 2.5.2.3 Deep LAC

将局部定位，对齐，分类集合在一个神经网络中，最终在有 BBox 标注信息的情况下，在 CUB-Bird 数据集上获得了 80.3% 的准确率。

### 2.5.2.4 其他算法

part-stack CNN：采用 FCN 进行局部定位，并用高斯核对特征图去噪，最终在 CUB-Bird 数据集上，在用了 BBox (Bounding Box) 和 Parts (Parts Annotation) 标注信息时获得了 76.2% 的准确率。

Multi-proposal consensus：利用改进的 AlexNet 提取关键点和区域，并用 SVM 分类，并用到了 BBox 标注信息，在 CUB-Bird 数据集上获得了 80.3% 的准确率。

Coarse-to-fine：同时整合多个卷几层特征，由粗到细进行分类，在使用 BBox 标注信息的情况下，在 CUB-Bird 数据集上获得 83.7% 的准确率。

## 2.5.3 弱监督算法

### 2.5.3.1 空间转换网络

用空间转换器完成对前景图像的对齐操作，在 CUB-Bird 数据集上获得 84.1% 的准确率。

### 2.5.3.2 B-CNN

在只有标签的情况下，在 CUB-Bird 数据集上分类准确率最高：84.1%。

### 2.5.3.3 Constellations

利用了卷积神经网络本身产生的关键点来提取局部区域的信息，并在 CUB-Bird 数据集上获得了较高的准确率：81%。

### 2.5.3.4 其他方法

两级注意力算法：关注对象级和局部级的特征，并对特征进行聚类，在 CUB-Bird 数据集上获得了 77.9% 的准确率。

FCAN：用全卷积网络来获得局部信息，也使用了注意力机制，在 CUB-Bird 数据集上获得较高的准确率 82%，而且对比了有 boundingbox 的结果 83.3%。

FCN attention：使用 FCN 网络来定位局部特征，在 CUB-Bird 数据集上获得了 82% 的准确率。

DVAN：同时提取多个重要区域的特征取平均值，在 CUB-Bird 数据集上获得了 79% 的准确率。

Subset Feature Learning：用聚类的方法将视觉上相近的图片聚集到一起，形成 K 个子集，再用 CNN 进行分类，在 CUB-Bird 数据集上分类准确率为 77.5%。

Attention for fine-grained categorization: 将 RNN 与 GoogLeNet 结合, 在 Stanford Dogs 数据集上得到了 76.8% 的准确率。

CNN tree: 形成树状 CNN 网络, 每次将“干扰项”作为新的子集, 重新训练。

## 2.6 细粒度分类存在的问题和挑战

# Chapter 3

---

## 实验

---

现已提出的细粒度分类的方法, 主要的关注的是以下几个方面: 1、针对数据集不大, 图片质量不高的数据预处理和模型微调方法; 2、针对细粒度分类的有效的局部特征提取的方法; 3、对于提取到的特征的处理方法, 包括特征选择, 降维, 特征表示及特征融合。

### 3.1 数据增强和模型微调的方法:

#### 3.1.1 Web Data

Web Data[15][强监督] 方法通过开发网络照片来对 CUB-Bird 数据集进行扩充, 从而提高了分类准确率, 这可作为数据增强的一个思路。

#### 3.1.2 Webly-supervised

Webly-supervised[16] [强监督] 主要和 Web Data 方法类似, 都用到了网页中的图片。该方法将数据集中带有 BoundingBox 信息和 Part Landmarks 信息的图片和网页上的没有这些信息的图片很好的结合, 提高了准确率。这也是数据增强的一个思路。

#### 3.1.3 Deep Metric Learning

DML[1][弱监督] 方法中介绍了建立数据集的方法, 通过 Google 搜索或在一些社交网站上找到后, 他们提出了 Bootstrapping 框架来处理这些数据。该数据获取及处理的方法值得学习, 但是本文提出的流形学习模式可先不过多关注。

### 3.2 局部特征的获取:

#### 3.2.1 Deep LAC

Deep LAC[4][强监督] 将定位, 对齐和分类集合到一起, 重点考虑 head 和 body 的局部特征, 但是他需要 BoundingBox 标注信息, 而且在有 BoundingBox 标注信息时, 准确率为 80.3%, 可先不考虑实验。

### 3.2.2 partRCNN

Part-based R-CNNs[17][强监督] 主要是利用 R-CNN 算法进行对象和该对象区域的检测, 该方法在使用 BoundingBox 和 Part 信息的情况下, 获得了 73.9% 的准确率。R-CNN 是局部检测的经典方法, 但是分类准确率不高, 而且需要的人工标注信息多, 可先不考虑实验, 但理论要研究清楚。

### 3.2.3 B-CNN

B-CNN[5][弱监督] 方法采用两个 subnet 来分别提取不同的特征, 并对特征进行融合, 该方法中的子网络的结合方式是准确率提高的重要原因, 而且该方法在只有类别表现的情况下, 在 CUB-Bird 数据集上获得了 84.1% 的准确率, 该方法需要实验。

### 3.2.4 CNN-tree

CNN-tree[14] [弱监督] 将每次分类后的干扰项挑选出来, 打包成不同的子集再重新分类, 如此反复, 从而形成了树状结构。该方法以 top-1 error 和 top-5 error 为评价指标, 最终在以 AlexNet 为基准, 在 ILSVRC2014 数据集上获得了 37.19% 的 top-1 error 和 16.23% 的 top-5 error。此方法网络结构较为复杂, 而且性能提升有限。

### 3.2.5 FCAN

FCAN[6][弱监督] 方法用全卷积网络来做特征提取, 用 softmax 做分类, 在 CUB-Bird 上有 BBox 信息时得到了 84.3% 的准确率, 在仅用标签信息时获得了 82% 的准确率。此方法还考虑到了注意力机制, 值得关注。

### 3.2.6 Multi-proposal Consensus

Multi-proposal Consensus[10][弱监督] 主要在于局部定位和特征提取, 并移除掉能见度差的特征点, 最后用 SVM 进行分类, 在 CUB-Bird 的准确率为 78.3%。分类效果一般。

### 3.2.7 Part-stack CNN

Part-stack CNN[3][强监督] 采用 FCN 网络进行局部定位, 用卷积神经网络提取特征并用高斯核对得到的特征图去噪, 最后采用基于对象级和局部级的两通道卷积神经网络进行分类。最终在有 BBox 标注信息情况下, 在 CUB-Bird 数据集上获得了 76.6% 的分类准确率。其局部定位方法值得看, 但是最终分类效果一般。

### 3.2.8 Pose Normalization

Pose Normalization[11][强监督] 采用改进的聚类算法来对不同层次的局部信息进行姿态对齐操作。在有 part 和 BBox 标注信息的情况下, 在 CUB-Bird 数据集上获得了 85.4% 的准确率。此方法也是将 SVM 和 CNN 结合的很好的例子。

### 3.2.9 Constellations

星座算法 [8][弱监督] 主要是用到星座算法对卷积网络提取到的特征中的关键点进行筛选后提取出局部区域, 并最终在仅有标签信息的情况下, 得到了 81% 的准确率。

### 3.2.10 Two-level Attention

Two-level Attention[12][弱监督] 模型提出对局部级和对象级分类的方法, 但是在只有标签信息的情况下在 CUB-Bird 数据集上仅得到 69.7% 的分类准确率。

## 3.3 特征处理:

### 3.3.1 mixDCNN

mixDCNN[2][强监督] 的方法是将各 CNN 网络结合, 即将卷积神经网络提取到的特征结合。比如 AlexNet 是将第一层全连接层之后的输出进行聚类, GoogLeNet 是将分类前的最后一层进行聚类, 从而实现特征的融合。虽然该方法也用到了 BoundingBox 信息, 最高准确率为 81.2%, 但是这种特征聚类的方法值得一试。

### 3.3.2 Part-Based Image Representation

Part-Based Image Representation[18][弱监督] 方法从整张图片中提取局部特征, 再从局部特征中选择有用的局部特征, 然后将这些特征匹配到全局中进行分类。此方法虽然只获得了 78.92% 的准确率, 但是这种在深度神经网络中直接进行特征选择的方式值得学习, 而且作者也给出了选择比率不同时的效果。

### 3.3.3 Coarse-to-Fine

Coarse-to-Fine[13] [强监督] 是由粗到细的细粒度分类方法, 该方法是为了解决最后层卷积层输出的特征图像全局化而忽视了许多细节特征因而不适用细粒度分类问题而提出的, 他们将不同卷积层的输出相结合, 从而关注到细节特征, 在有 BBox 信息的情况下, 在 CUB-Bird 数据集上获得了 83.7% 的准确率。此文发表在 2016ICIP 上, 难度适宜, 方法也比较新颖, 但是链接结构可能并不是很有效, 而且需要 BBox 标注信息, 此方法值得一试。

### 3.3.4 DVAN

DVAN[19][弱监督] 用到了 LSTM 来学习 attentiveness and discrimination. 并融合多个区域识别的特征, 最终取所有预测结果的平均值, 在 CUB-Bird 上获得 79% 的准确率。

### 3.3.5 Filter-Specific

Filter-Specific[7][弱监督] 在最后一层卷积层后面加上了滤波器来处理冗余问题。并且使用了玉米分类的新的数据集, 但是他们提出的滤波方法并不是很有效, 所以待定。

### 3.3.6 Transformer Networks

空间转换网络 [9][弱监督] 提出了可以嵌入到卷积神经网络中的空间转换模型, 通过该模型可以使得 feature map 具有空间转换能力。该方法不仅在 CUB-Bird 数据集上实验, 也应用于 MNIST 数据集, 最终在仅有标签信息的情况下, 在 CUB-Bird 数据集上获得了 84.1% 的准确率。此方法不仅仅适用于细粒度分类问题, 对于许多分类数据集都有适用的可能性。

## 3.4 其他的改进方法:

---

## 参考文献

---

- [1] Yin Cui, Feng Zhou, Yuanqing Lin, and Serge Belongie. Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1153–1162, 2016.
- [2] ZongYuan Ge, Alex Bewley, Christopher McCool, Peter Corke, Ben Upcroft, and Conrad Sanderson. Fine-grained classification via mixture of deep convolutional neural networks. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–6, 2016.
- [3] Shaoli Huang, Zhe Xu, Dacheng Tao, and Ya Zhang. Part-stacked cnn for fine-grained visual categorization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [4] Di Lin, Xiaoyong Shen, Cewu Lu, and Jiaya Jia. Deep lac: Deep localization, alignment and classification for fine-grained recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1666–1674, 2015.
- [5] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear cnn models for fine-grained visual recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1449–1457, 2015.
- [6] Xiao Liu, Tian Xia, Jiang Wang, Yi Yang, Feng Zhou, and Yuanqing Lin. Fine-grained recognition with automatic and efficient part attention.
- [7] Hao Lu, Zhiguo Cao, Yang Xiao, Zhiwen Fang, and Yanjun Zhu. Fine-grained maize cultivar identification using filter-specific convolutional activations. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 3718–3722. IEEE, 2016.
- [8] Simon Marcel and Erik Rodner. Neural activation constellations: Unsupervised part model discovery with convolutional networks. *ICCV*, pages 1143–1151, 2015.



- [9] Jaderberg Max, Simonyan Karen, Zisserman Andrew, and kavukcuoglu koray. Spatial transformer networks. *NIPS*, 2015.
- [10] Kevin J Shih, Arun Mallya, Saurabh Singh, and Derek Hoiem. Part localization using multi-proposal consensus for fine-grained categorization. *arXiv preprint arXiv:1507.06332*, 2015.
- [11] Branson Steve, Horn Grant, Van, Belongie Serge, and Perona Pietro. Bird species categorization using pose normalized deep convolutional nets. *CVPR*, 2014.
- [12] Xiao Tianjun, Xu Yichong, Yang Kuiyuan, Zhang Jiaying, Peng Yuxin, and Zhang Zheng. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. *CVPR*, pages 842–850, 2015.
- [13] Yongzhong Wang, Xu-Yao Zhang, Yanming Zhang, Xinwen Hou, and Cheng-Lin Liu. Exploiting coarse-to-fine mechanism for fine-grained recognition. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 649–653. IEEE, 2016.
- [14] Zhenhua Wang, Xingxing Wang, and Gang Wang. Learning fine-grained features via a cnn tree for large-scale classification. *arXiv preprint arXiv:1511.04534*, 2015.
- [15] Zhe Xu, Shaoli Huang, Ya Zhang, and Dacheng Tao. Augmenting strong supervision using web data for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2524–2532, 2015.
- [16] Zhe Xu, Shaoli Huang, Ya Zhang, and Dacheng Tao. Webly-supervised fine-grained visual categorization via deep domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
- [17] Ning Zhang, Jeff Donahue, Ross Girshick, and Trevor Darrell. Part-based r-cnns for fine-grained category detection. In *European conference on computer vision*, pages 834–849, 2014.
- [18] Yu Zhang, Xiu-Shen Wei, Jianxin Wu, Jianfei Cai, Jiangbo Lu, Viet-Anh Nguyen, and Minh N Do. Weakly supervised fine-grained categorization with part-based image representation. *IEEE Transactions on Image Processing*, pages 1713–1725, 2016.
- [19] Bo Zhao, Xiao Wu, Jiashi Feng, Qiang Peng, and Shuicheng Yan. Diversified visual attention networks for fine-grained object classification. *CoRR*, abs/1606.08572, 2016.