1、Large Marge Softmax Loss

ICML2016 提出的 Large Marge Softmax Loss（L-softmax）通过在传统的 softmax loss 公式中添加参数 m，加大了学习的难度，逼迫模型不断学习更具区分性的特征，从而使得类间距离更大，类内距离更小。核心内容可以看下图：

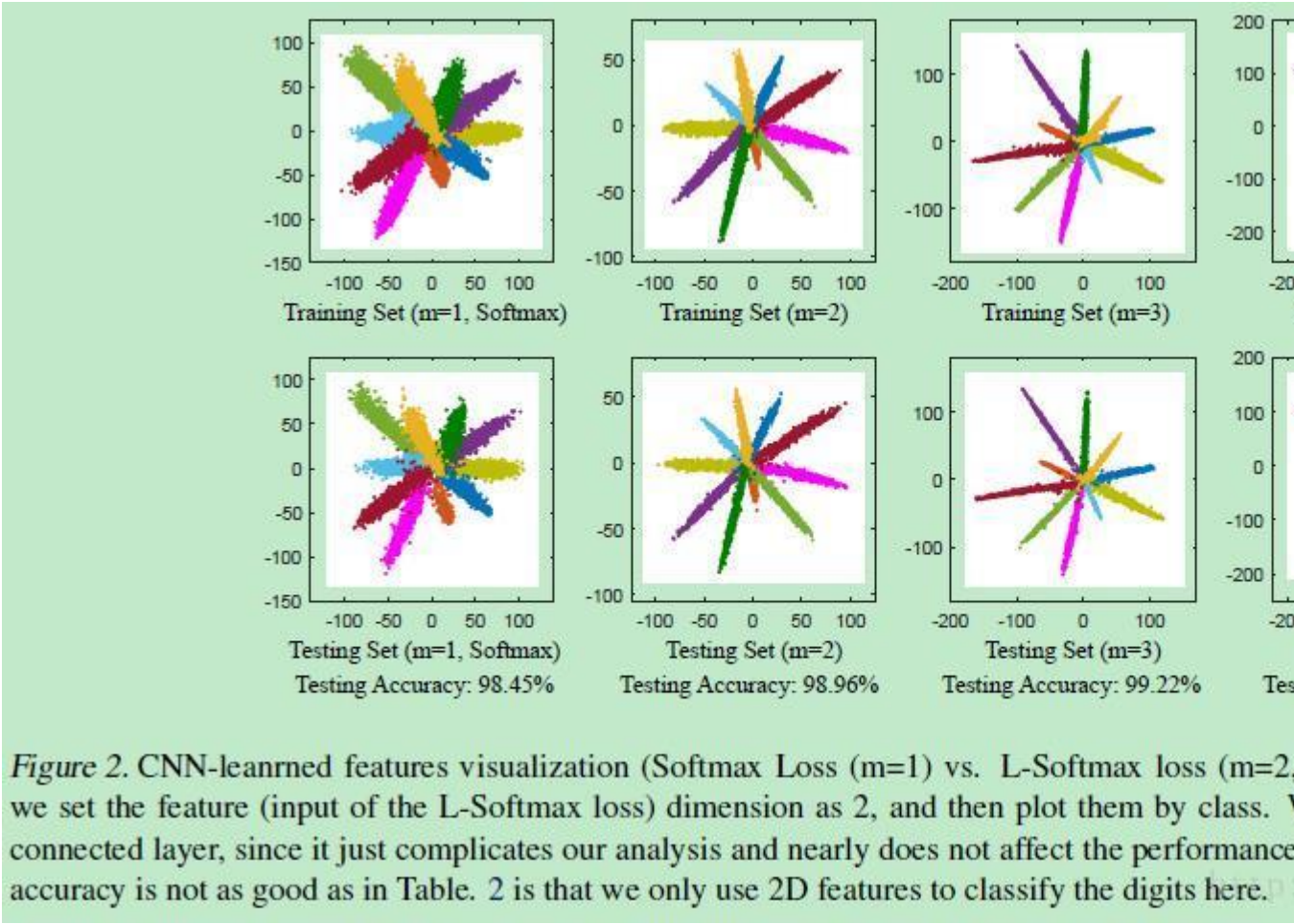


Figure 2. CNN-leanrned features visualization (Softmax Loss (m=1) vs. L-Softmax loss (m=2, we set the feature (input of the L-Softmax loss) dimension as 2, and then plot them by class. connected layer, since it just complicates our analysis and nearly does not affect the performance accuracy is not as good as in Table. 2 is that we only use 2D features to classify the digits here.

详细了解可以参看：损失函数改进之 Large-Margin Softmax Loss

2、Center Loss

ECCV2016 提出的 center loss 是通过将特征和特征中心的距离和 softmax loss 一同作为损失函数，使得类内距离更小，有点 L1，L2 正则化的意思。核心内容如下图所示：

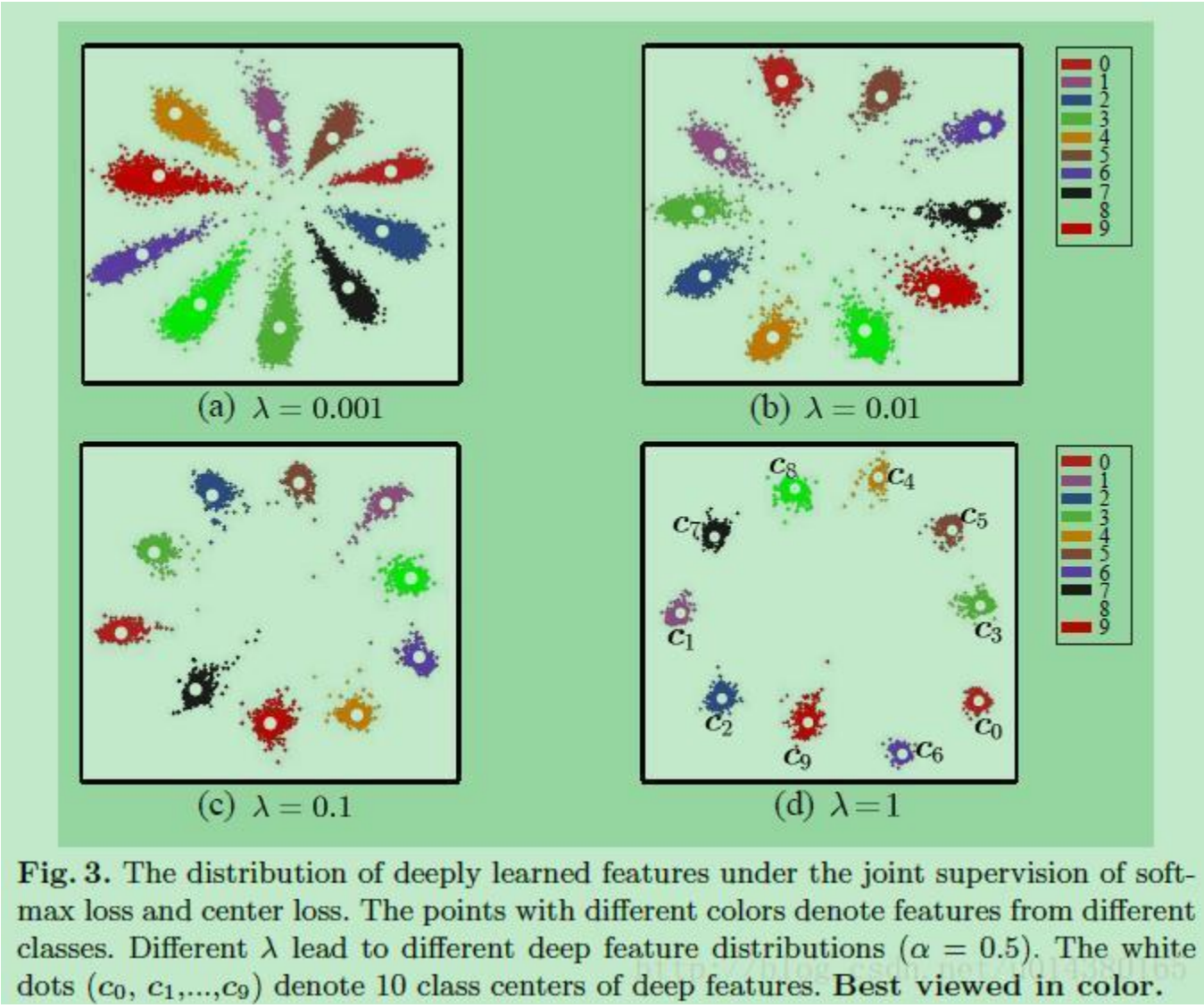
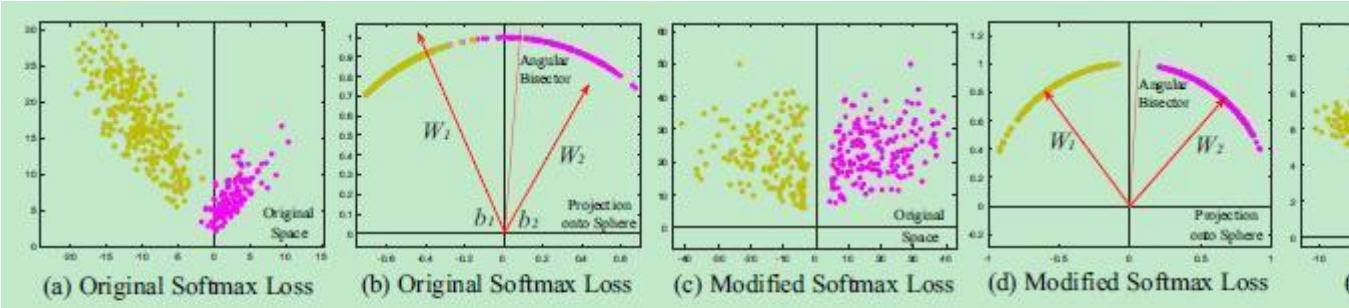


Fig. 3. The distribution of deeply learned features under the joint supervision of softmax loss and center loss. The points with different colors denote features from different classes. Different λ lead to different deep feature distributions ($\alpha = 0.5$). The white dots (c_0, c_1,...,c_9) denote 10 class centers of deep features. **Best viewed in color.**

详细了解可以参看：损失函数改进之 Center Loss

3、Angular Softmax Loss

CVPR2017 提出的 A-softmax loss（angular softmax loss）用来改进原来的 softmax loss，并将 A-softmax loss 用于人脸识别，这就是 SphereFace，作者还是 Large margin softmax loss 的原班人马。A-softmax loss 简单讲就是在 large margin softmax loss 的基础上添加了两个限制条件||W||=1 和 b=0，使得预测仅取决于 W 和 x 之间的角度。核心思想可以参看下面这个图。



(a) Original Softmax Loss (b) Original Softmax Loss (c) Modified Softmax Loss (d) Modified Softmax Loss

Figure 2: Comparison among softmax loss, modified softmax loss and A-Softmax loss. In this toy experiment, subset of the CASIA face dataset. In specific, we set the output dimension of FC1 layer as 2 and visualize t first class face features, while purple dots represent the second class face features. One can see that features l classified simply via angles, while modified softmax loss can. Our A-Softmax loss can further increase the ar

详细了解可以参看：SphereFace 算法详解

4、Focal Loss

Focal loss 是 Facebook 的 RBG 和 Kaiming 大神在 ICCV2017 的 best paper，主要是改进了目标检测（object detection）算法的效果，在 COCO 数据集上的 AP 和速度都有明显提升。核心思想在于概括了 object detection 算法中 proposal-free 一类算法准确率不高的原因在于：类别不均衡。于是在传统的交叉熵损失上进行修改得到 Focal Loss。

详细了解可以参看：Focal Loss

5、Additive Angular Margin Loss

这篇文章提出一种新的用于人脸识别的损失函数：additive angular margin loss，基于该损失函数训练得到人脸识别算法 ArcFace（或者叫 InsightFace）。ArcFace 的思想（additive angular margin）和 SphereFace（angular softmax loss）以及不久前的 CosineFace（additive cosine margin ）有一定的共同点，重点在于：在 ArchFace 中是直接在角度空间（angular space）中最大化分类界限，而 CosineFace 是在余弦空间中最大化分类界限，这也是为什么这篇文章叫 ArcFace 的原因，因为 arc 含义和 angular 一样。

公式如下：

$$L_7 = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}}, \tag{9}$$

subject to

$$W_j = \frac{W_j}{\|W_j\|}, x_i = \frac{x_i}{\|x_i\|}, \cos\theta_j = W_j^T x_i. \tag{10}$$

空金字塔池化（Spatial Pyramid Pooling）[3]

空间金字塔池化可以把任何尺度的图像的卷积特征转化成相同维度，这不仅可以让 CNN 处理任意尺度的图像，还能避免 cropping 和 warping 操作，导致一些信息的丢失，具有非常

重要的意义。

一般的 CNN 都需要输入图像的大小是固定的，这是因为全连接层的输入需要固定输入维度，但在卷积操作是没有对图像尺度有限制，所有作者提出了空间金字塔池化，先让图像进行卷积操作，然后转化成维度相同的特征输入到全连接层，这个可以把 CNN 扩展到任意大小的图像。

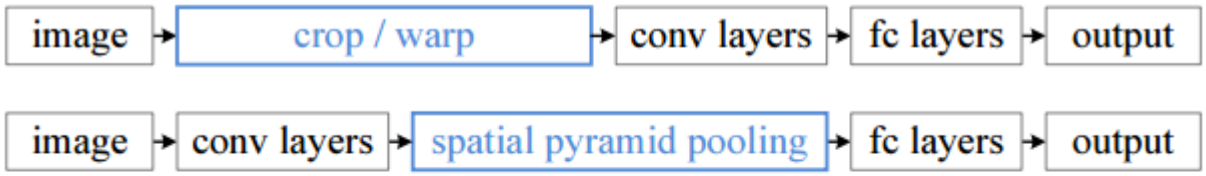
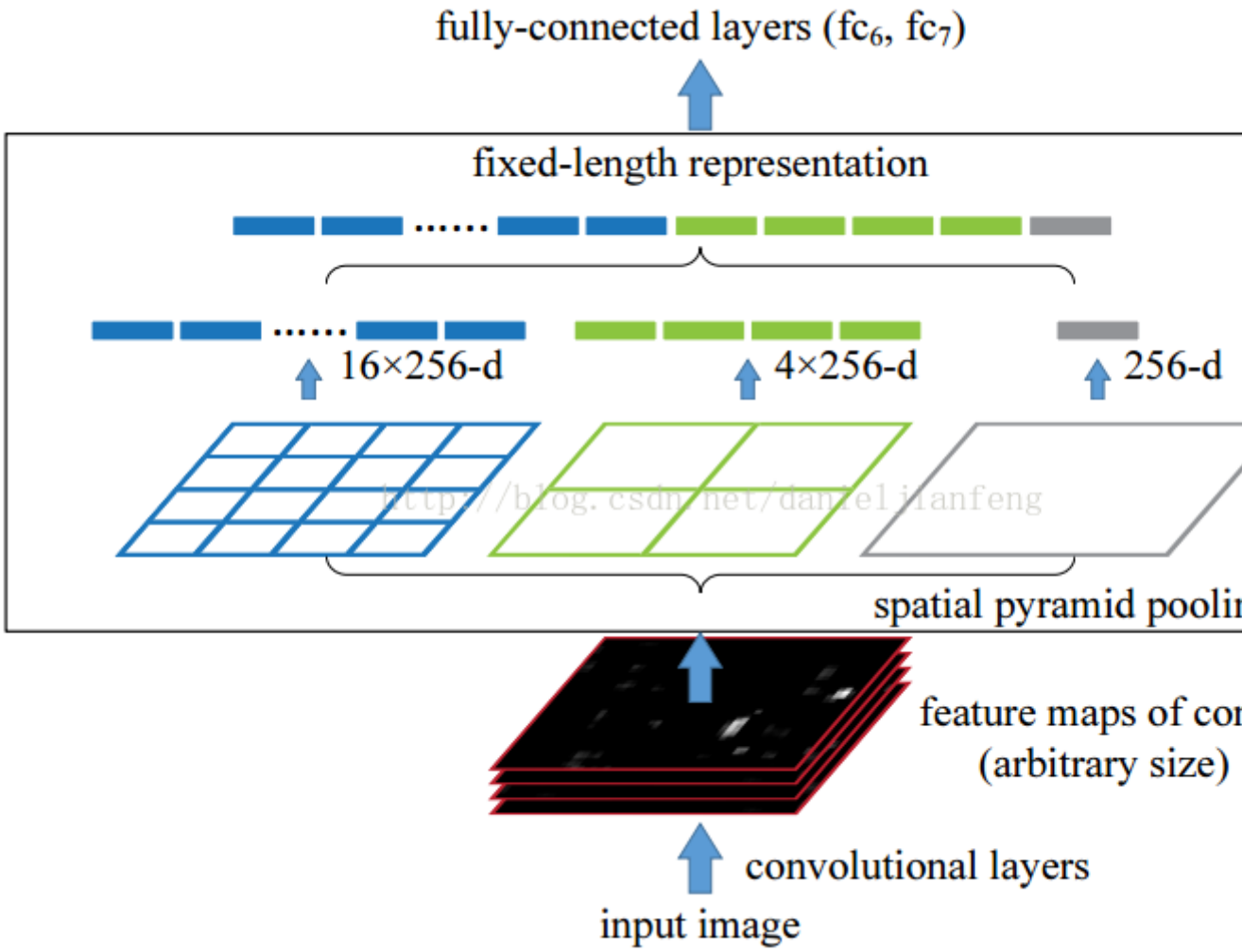


Figure 1: Top: cropping or warping to fit a fixed size. Middle: a conventional deep convolutional network structure. Bottom: our spatial pyramid pooling network structure.

空间金字塔池化的思想来自于 Spatial Pyramid Model，它一个 pooling 变成了多个 scale 的 pooling。用不同大小池化窗口作用于卷积特征，我们可以得到 1X1,2X2,4X4 的池化结果，由于 conv5 中共有 256 个过滤器，所以得到 1 个 256 维的特征，4 个 256 个特征，以及 16 个 256 维的特征，然后把这 21 个 256 维特征链接起来输入全连接层，通过这种方式把不同大小的图像转化成相同维度的特征。



对于不同的图像要得到相同大小的 pooling 结果，就需要根据图像的大小动态的计算池化窗口的大小和步长。假设 conv5 输出的大小为 a*a，需要得到 n*n 大小的池化结果，可以让窗口大小 sizeX 为，步长为 。下图以 conv5 输出的大小为 13*13 为例。

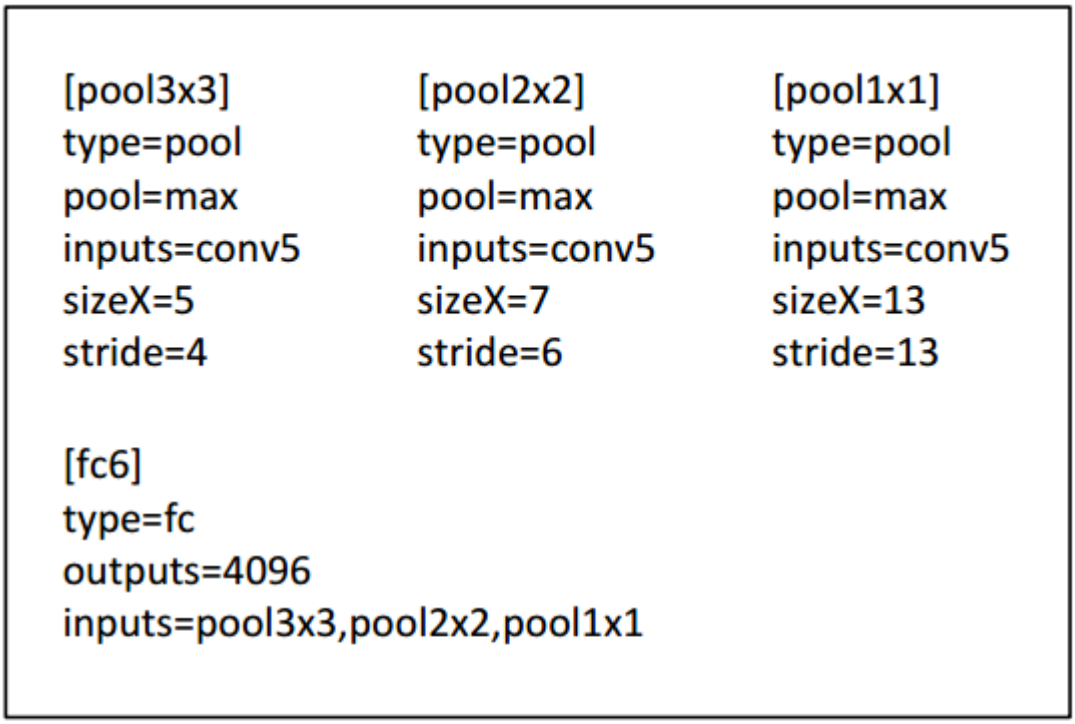


Figure 4: An example 3-level pyramid pooling in the convnet style [3]. Here sizeX is the size of the pooling window. This configuration is for a network whose feature map size of $conv_5$ is 13×13, so the $pool_{3\times3}$, $pool_{2\times2}$, and $pool_{1\times1}$ layers will have 3×3, 2×2, and 1×1 bins respectively.

1、Color Jittering：对颜色的数据增强：图像亮度、饱和度、对比度变化（此处对色彩抖动的理解不知是否得当）；
2、PCA Jittering：首先按照 RGB 三个颜色通道计算均值和标准差，再在整个训练集上计算协方差矩阵，进行特征分解，得到特征向量和特征值，用来做 PCA Jittering；
3、Random Scale：尺度变换；
4、Random Crop：采用随机图像差值方式，对图像进行裁剪、缩放；包括 Scale Jittering 方法（VGG 及 ResNet 模型使用）或者尺度和长宽比增强变换；
5、Horizontal/Vertical Flip：水平/垂直翻转；
6、Shift：平移变换；
7、Rotation/Reflection：旋转/仿射变换；
8、Noise：高斯噪声、模糊处理；

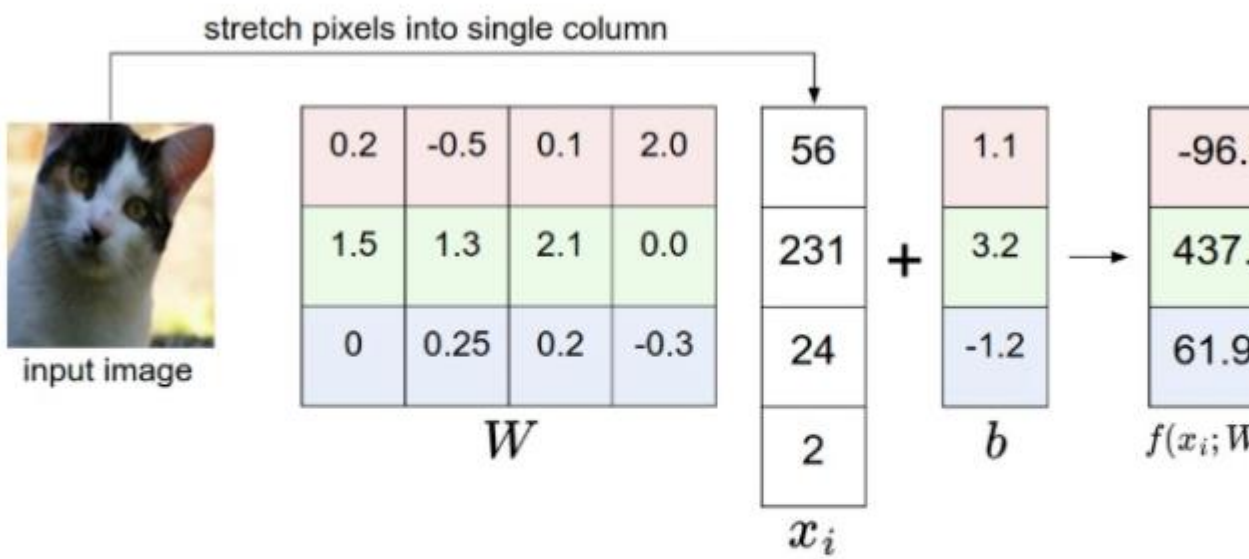
9、Label shuffle：类别不平衡数据的增广，参见海康威视 ILSVRC2016 的 report；另外，文中提出了一种 Supervised Data Augmentation

图像分类的主要过程包括图像预处理、特征提取和分类器设计。图像预处理包括图像滤波，如中值滤波[7]、均值滤波[8]、高斯滤波[9]以及图像归一化等操作，其主要作用是过滤图像中的一些无关信息，在简化数据的前提下最大限度地保留有用信息，增强特征提取的可靠性。特征提取是图像分类任务中最为关键的一部分，其将输入图像按照一定的规则变换生成另一种具有某些特性的特征表示，新的特征往往具有低维度、低冗余、低噪声、结构化等优点，从而降低了对分类器复杂度的要求，提高了模型性能。最后通过训练分类器对提取的特征进行分类，从而实现图像的分类。

传统的图像分类研究中，多数为基于图像特征的分类，即根据不同类别图像的差异，利用图像处理算法提取相应的经过定性或定量表达的特征，对这些特征进行数学统计分析或使用分类器输出分类结果。在特征提取方面，主要包括纹理、颜色、形状等底层视觉特征，尺度不变特征变换[10]、局部二值模式[11]、方向梯度直方图[12]等局部不变性特征，这些人工设计特征缺乏良好的泛化性能，且依赖于设计者的先验知识和对分类任务的认知理解。目前，海量、高维的数据也使得人工设计特征的难度呈指数级增加。
在分类器方面，主要包括 k NN(k-nearest neighbor,k 最近邻）决策树[14]、SVM(support vector machine，支持向量机）[15]、人工神经网络[16]等方法。这些分类器大大地提升了图像分类的效果，但对于处理庞大的图像数据、图像干扰严重等问题，其分类精度无法满足实际需求，故传统分类器不适合复杂图像的分类。
深度学习[17]是机器学习的一种新兴算法，因其在图像特征学习方面具有显著效果而受到研究者们的广泛关注。相较于传统的图像分类方法，其不需要对目标图像进行人工特征描述和提取，而是通过神经网络自主地从训练样本中学习特征，提取出更高维、抽象的特征，并且这些特征与分类器关系紧密，很好地解决了人工提取特征和分类器选择的难题，是一种端到端的模型。

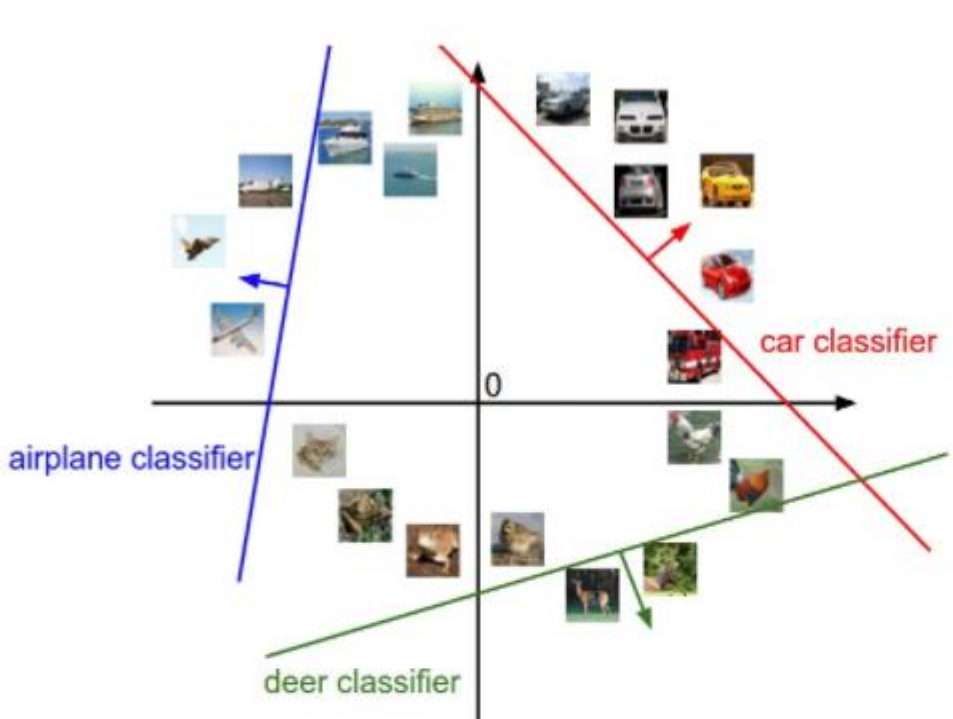
传统方法 3 步：图像预处理、特征提取和分类器设计，每一种都有多种方法。
深度学习方法 2 步：图像预处理和图像识别，其中图像识别中的骨干网络实现特征提取的功能，后面的 softmax 等实现图像分类。

2.3.1.1 线性分类解释



例子：为了方便看见，假设图像只有4个像素（4个单色像素，这里为了简单，不考虑3通道
们有3个类（猫（CAT）、狗（DOG）、轮船（SHIP）。我们将图像像素拉伸成一列，然后
到每个类的分数。$y = Wx + b = [3,4] * [4,1] + [3,1] = [3,1]$。上图图中的权重计算结果结果
给我们的猫图像分配一个非常低的猫分数。得出的结果偏向于狗。

如果可视化分类，我们为了方便，将一个图片理解成一个二维的点，在下面坐标中显示如下



- 解释：w的每一行都是其中一个类的分类器。这些数字的几何解释是，当我们改变w的
中相应的线将以不同的方向旋转。而其中的偏置是为了让我们避免所有的分类器都过原

- 总结：分类器的权重矩阵其实是对应分类的经过训练得到的一个分类模板，通过测试数