根据学校课堂纪律的要求

请同学们坐在前五排

数字媒体技术基础

Meng Yang

www.smartllv.com

SUN YAT-SEN University

机器智能与先进计算教育部重点实验室

智能视觉语言学习研究组

Course Outline

- 8. 数字媒体分类技术
- 8.1 传统分类算法
  - 8.1.1 K-近邻分类器
  - 8.1.2 支持向量机
  - 8.1.3 稀疏协同表示分类器
- 8.2 图像识别任务
- 8.3 语音识别任务
- 8.4 文本分类任务

第二部分

8.2 图像识别任务

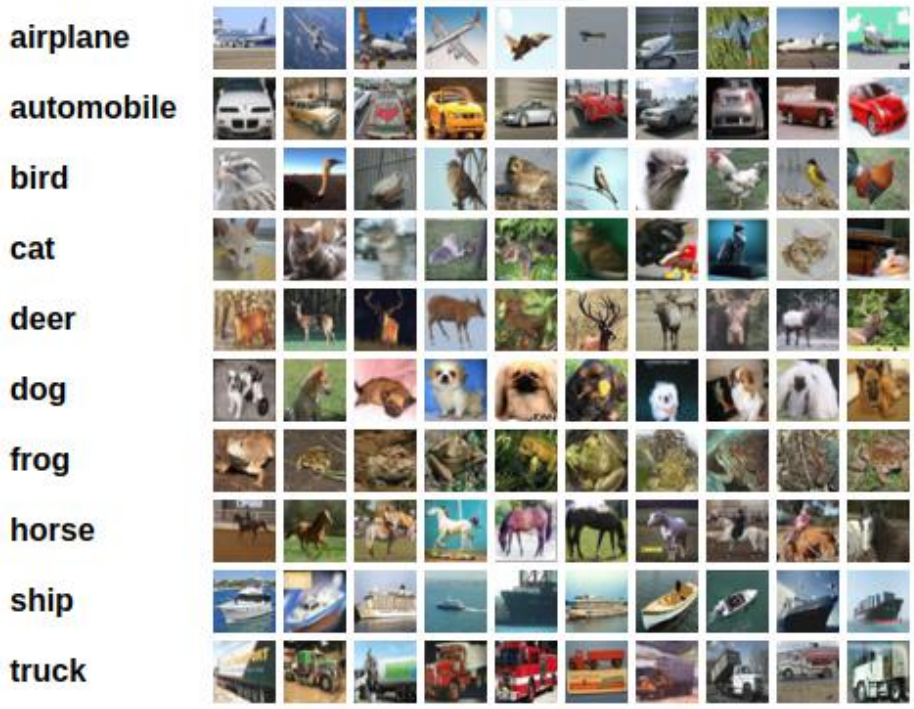
8.2 图像识别任务

- 图像识别任务是计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术，是深度学习算法的一种实践应用。

- 图像识别任务具体可以分为：
  - 图片分类任务：图像是什么
  - 目标检测任务：图像中目标在哪里
  - 语义分割任务：从像素级别上回答上面两个问题

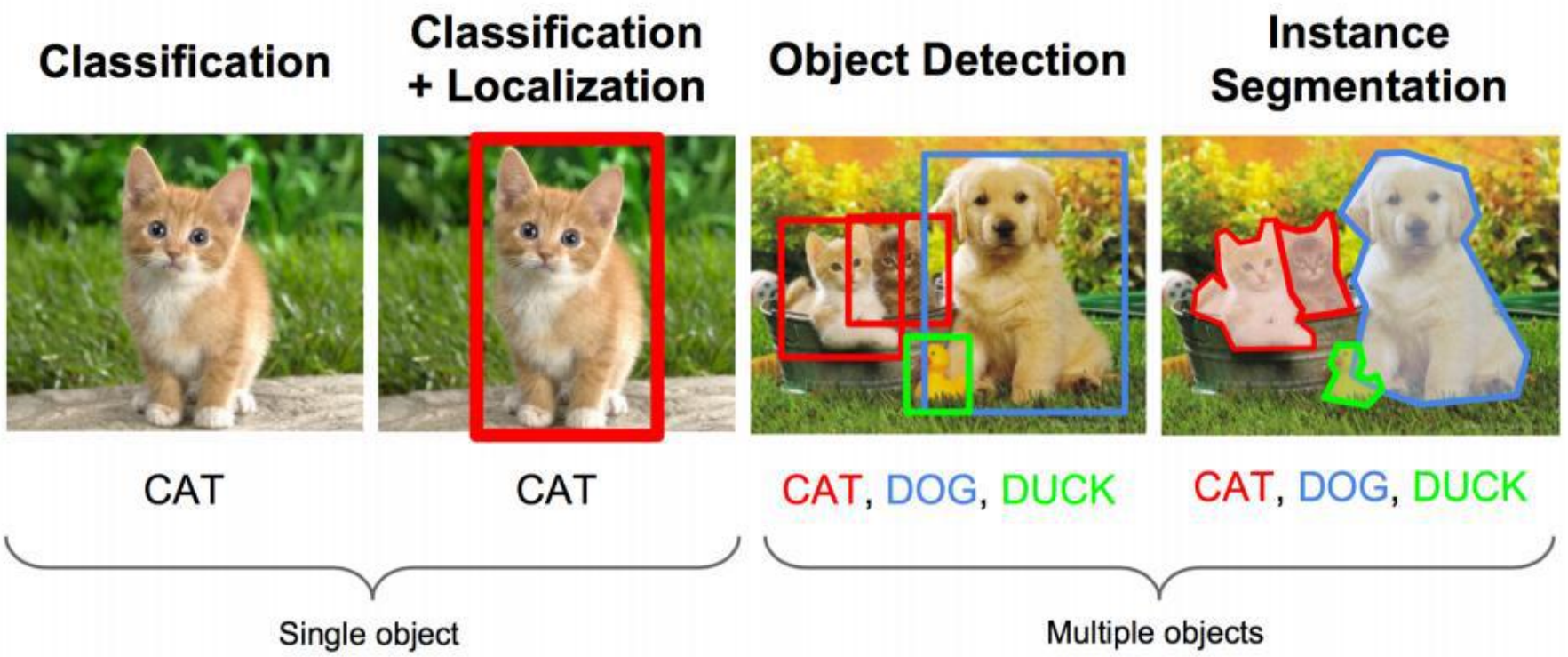
8.2 图像识别任务

- 图片分类任务：即通过常用的神经网络模型得到给定图片所属类别。小型数据集有以下两个：Cifar-10和mnist：



8.2 图像识别任务

- 目标检测任务：在图像分类的基础上，不仅需要判断图片中的物体，还要再图中标记出它的位置。



8.2 图像识别任务

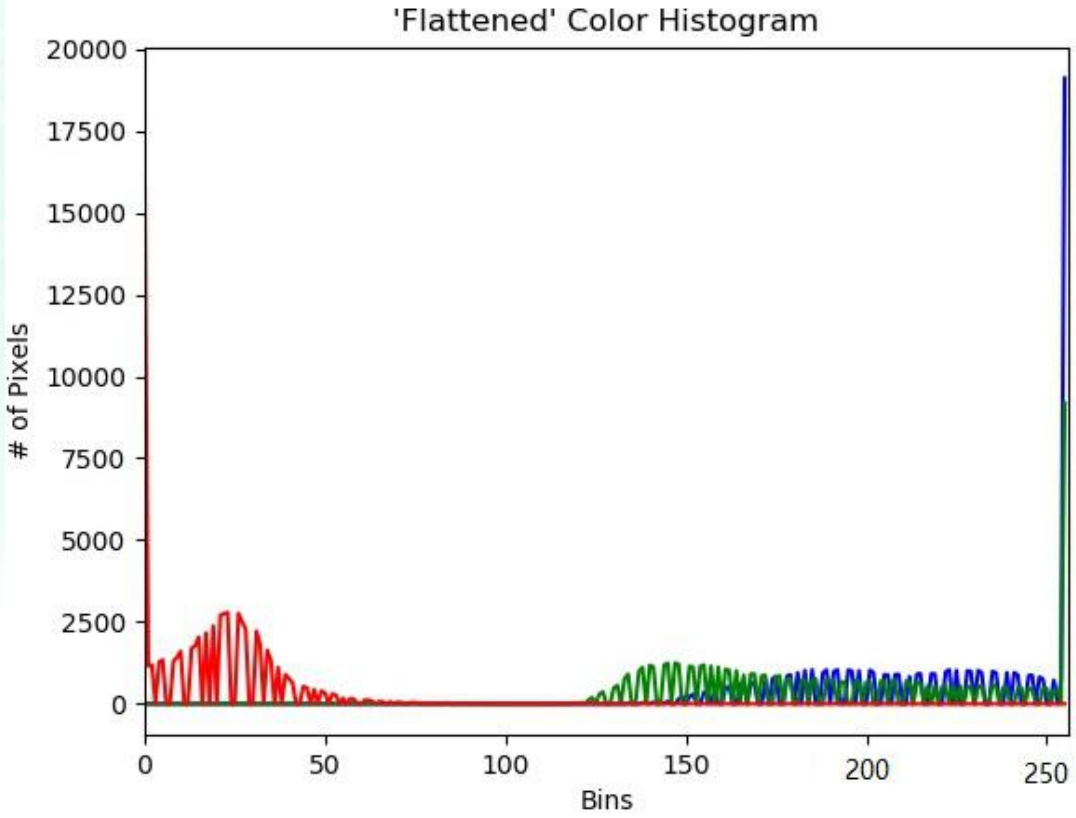
- 语义分割任务：是让计算机根据图像的语义进行分割。目标是从像素的角度分割出图片中的不同对象，对原图中的每个像素都进行标注。

8.2 图像识别任务

- 图像分类的一般流程：
  - 提取图像特征；
  - 模型处理特征；
  - 分类器得到最终类别概率。
- 图像特征提取经历了从机器学习到深度学习的发展历程，可以分为：
  - 图像颜色特征；
  - 图像BoW特征；
  - 深度特征。

8.2 图像识别任务

- 图像颜色特征：
  - 量化颜色直方图：将颜色空间量化，每块颜色由单元中心表示，统计落在量化单元上的像素的数量：



8.2 图像识别任务

问题？

- 图像颜色特征：
  - 颜色矩：简单有效的颜色特征表示方法，有一阶矩(均值)，二阶矩(标准差)和三阶矩(斜度)等，由于颜色信息主要分布在低阶矩中，因此到三阶矩足以表达图像颜色分布，颜色矩已证明可有效地表示图像中颜色分布。
  - 一阶矩表示为：$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{i,j}$
  - 二阶矩表示为：$\sigma_i = (\frac{1}{N}\sum_{j=1}^{N}(p_{i,j} - \mu_i)^2)^{\frac{1}{2}}$
  - 三阶矩表示为：$s_i = (\frac{1}{N}\sum_{j=1}^{N}(p_{i,j} - \mu_i)^3)^{\frac{1}{3}}$
  - 图像三个分量Y, U, V的前三阶颜色矩组成一个向量：

$$F_{color} = [\mu_Y, \sigma_Y, s_Y, \mu_U, \sigma_U, s_U, \mu_V, \sigma_V, s_V]$$

三阶矩：偏度，是统计数据分布偏斜方向和程度的度量，是统计数据分布非对称程度的数字特征。

8.2 图像识别任务

- 图像BoW特征：借鉴了文本特征的思想，文本由一系列的基本单元组成，通常基本单元为单词；一幅图像也可以看成是由一系列的基本单元组成，这些图像中的基本单元称为视觉单词。

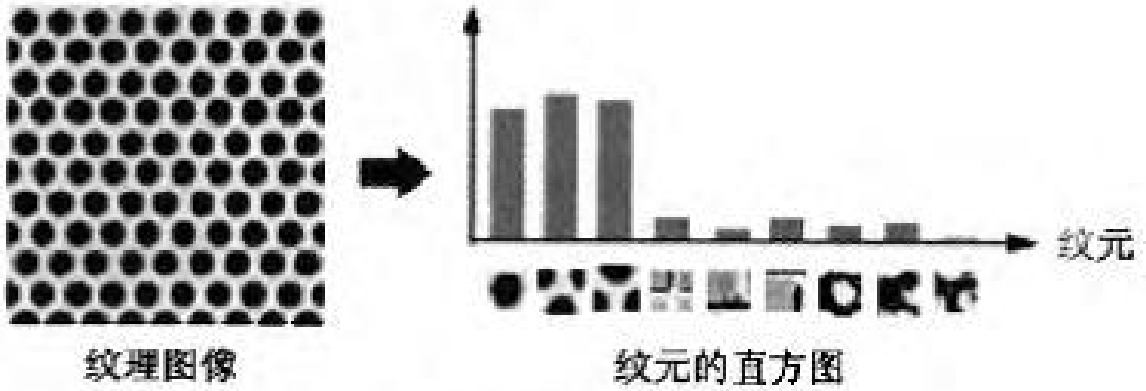


图 1　纹理图像用 Bag of Word 模型表示

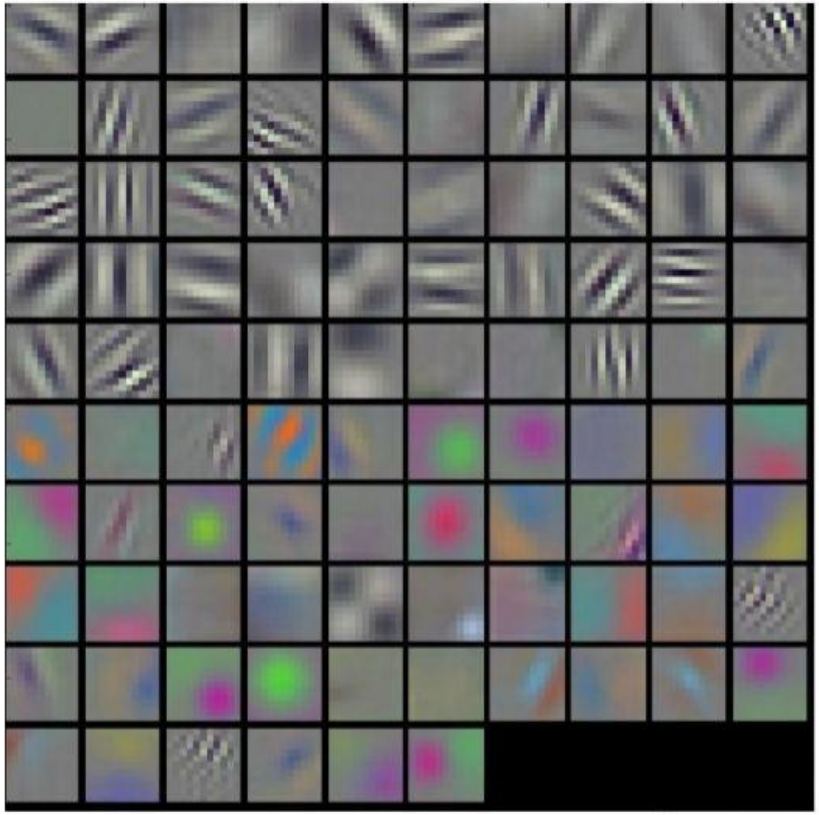
8.2 图像识别任务

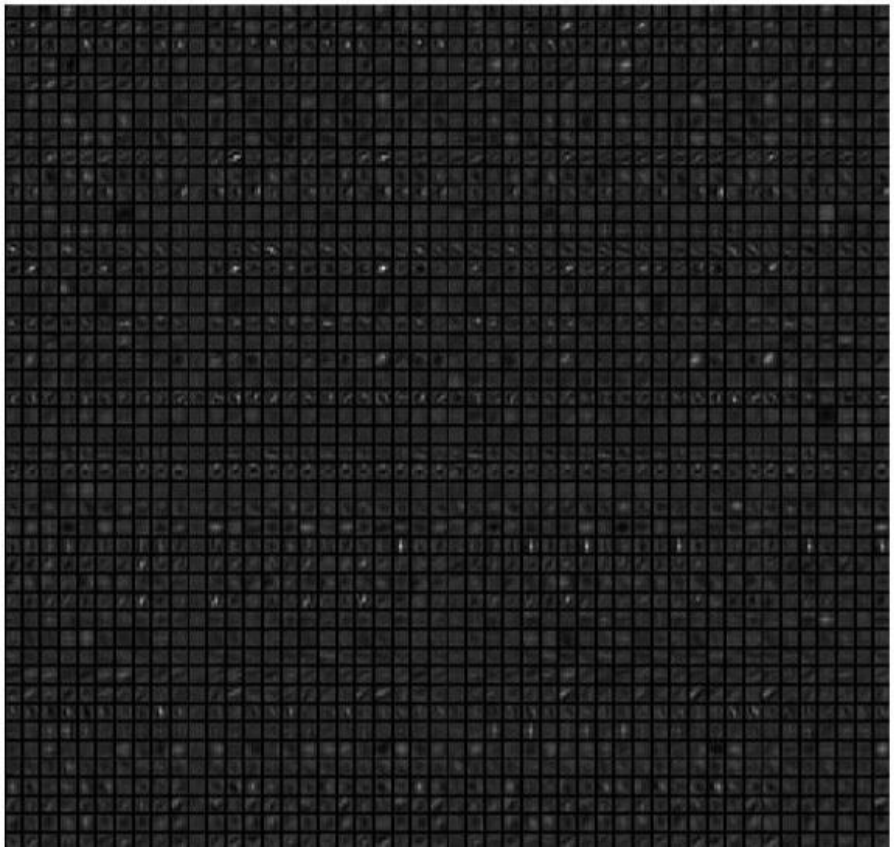
- 在2012年ImageNet挑战中，获奖模型为多伦多大学的AlexNet，他们强调他们的方法是“深度学习”，包括许多层叠在彼此智商的神经网络层和变换，包括卷积、全连接层、归一化层和最大池化层。

- 深度神经网络模型改进了图像特征提取的方法，模型从局部图像邻域中提取特征开始，后续层建立在前层的输出上。每一层提取的特征也有所区别，如AlexNet第一个卷积层提取出不同方向上灰度边缘和纹理；第二个卷积层是各种光滑图案检测器。

8.2 图像识别任务

- AlexNet第一个卷积层提取出不同方向上灰度边缘和纹理：
  AlexNet第二个卷积层是各种光滑图案检测器：





8.2 图像识别任务

- 图像分类任务得益于深度学习的发展，目前已经达到了很高的正确率。
- 人脸识别任务目前在LFW数据集中能达到99.6%以上的正确率(ArcFace)。
- 物体分类任务在cifar10数据集上正确率目前能达到80%以上。

8.2 图像识别任务

- 几个经典的图像识别任务深度模型：
  - LeNet-5：早期卷积神经网络中最有代表性的架构，用于手写数字十八别的卷积神经网络；
  - AlexNet：2012年ILSVRC冠军，6千万参数，自此之后，CNN称为图像识别分类的核心算法模型；
  - VGG：2014年ILSVRC亚军网络，1.38亿参数。由于网络结构十分简单，很适合迁移学习，至今VGG-16、VGG-19仍广泛使用；
  - ResNet：核心是带短连接的残差模块，其中主路径有两层卷积核（Res34），短连接把模块的输入信息直接和经过两次卷积之后的信息融合，有效增加了CNN的深度；
  - DenseNet、SENet ……

第三部分

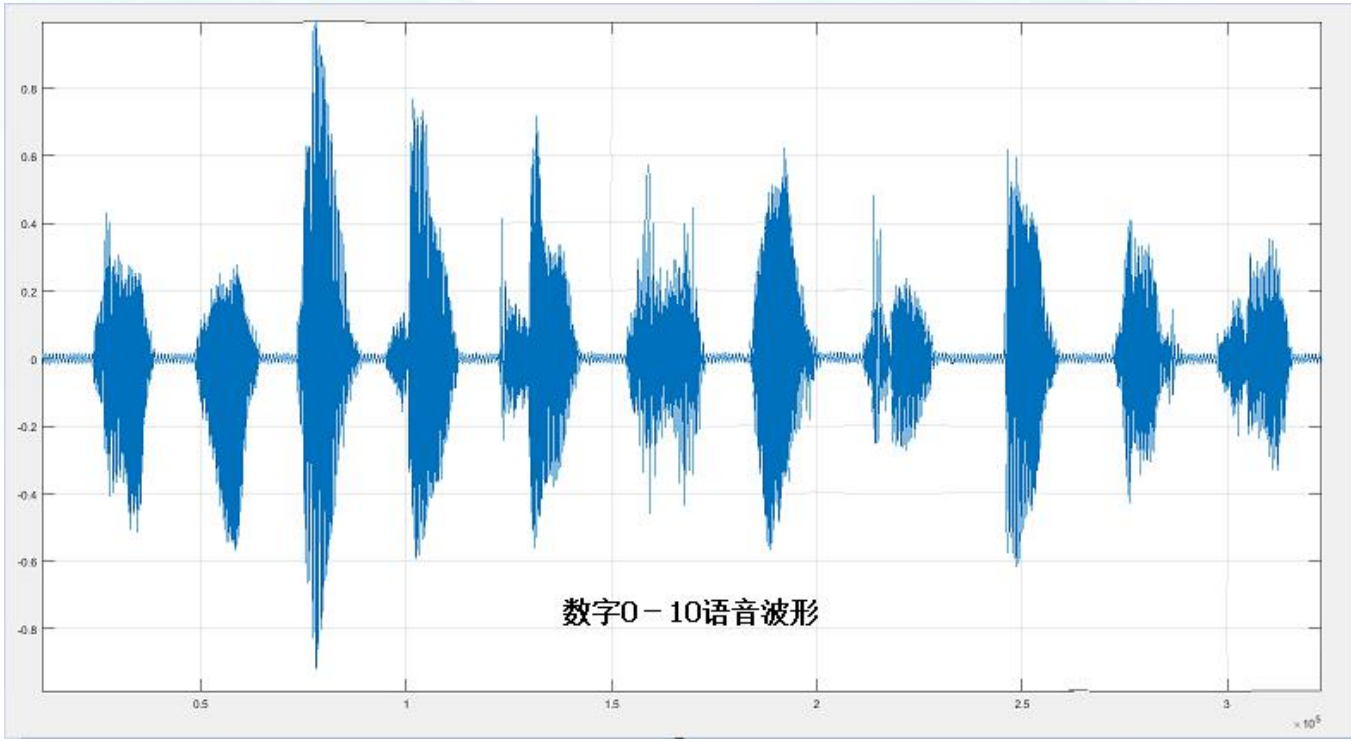
8.3 语音识别任务

8.2 语音识别任务

- 在人类的交流和知识传播中，大约70%的信息来自于语音。
- 语音识别是一门交叉学科，目的是与机器进行语音交流，让机器明白人们说什么。
- 语音识别系统的流程有：
  - 特征提取
  - 声学模型
  - CTC解码(Connectionist Temporal Classification是一种让网络自动学会对齐的好方法，十分适合语音识别和书写识别。)

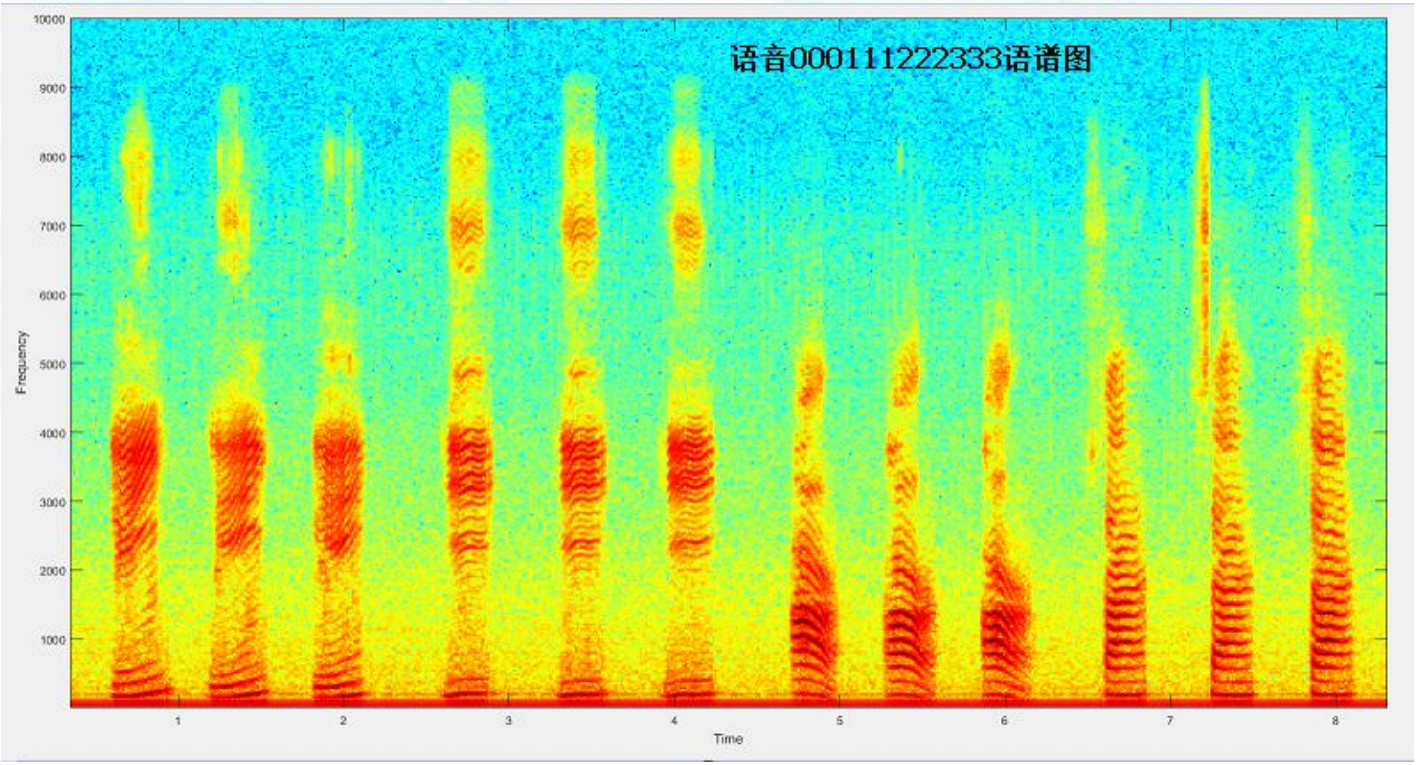
8.2 语音识别任务

- 特征提取：将普通的wav语音信号通过分帧加窗等操作转换为神经网络需要的二维频谱图像信号，即语谱图。
- 数字0-10语音波形如下：



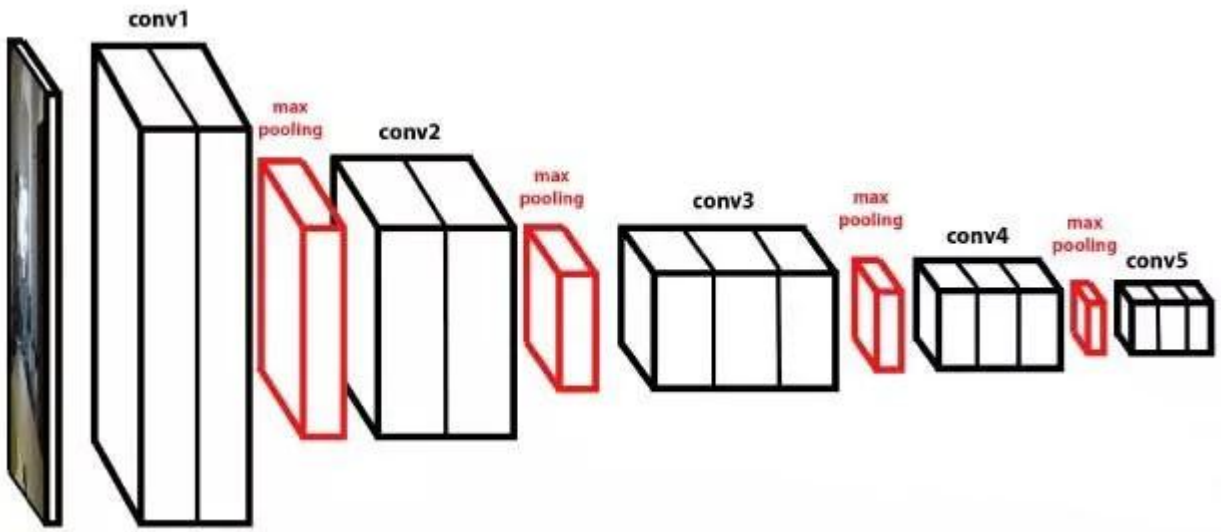
8.2 语音识别任务

- 语谱图：先将语音信号作傅里叶变换，然后以横轴为时间，纵轴为频率，用颜色表示幅值即可绘制出语谱图。在一幅图中表示信号的频率、幅度随时间的变化，故也称“时频图”。
- 语音000111222333的语谱图如下：



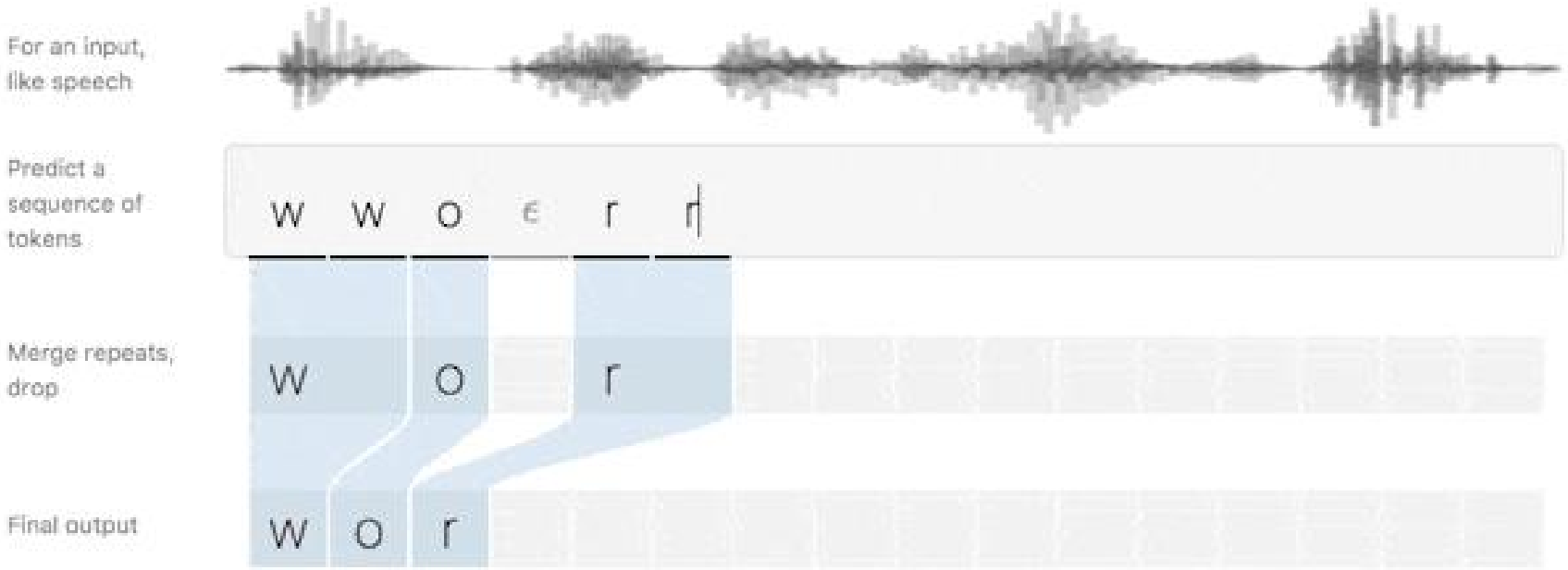
8.2 语音识别任务

- 声学模型：描述一种语言的基本单位被称为音素(Phoneme)。英语中有大约50多个音素，可以用单音素(monophone)模式表示单词的发音。但是在实际中很多发音是连在一起的，即音素构成音节，因此声学模型需要对不同语言，设置不同的多音素状态，从信号中的每一帧抽取不同的特征。比如参考VGG的深层卷积神经网络作为网络模型，并训练。



8.2 语音识别任务

- CTC解码：在语音识别系统的声学模型的输出中，往往包含了大量连续重复的符号，因此，我们需要将连续相同的符合合并为同一个符号，然后再去除静音分隔标记符，得到最终实际的语音拼音符号序列。



8.2 语音识别任务

- 语言模型：使用统计语言模型，将拼音转换为最终的识别文本并输出。
- 语言模型标准定义：对于语言序列$w_1, w_2, \ldots, w_n$，语言模型就是计算该序列的概率，即$P(w_1, w_2, \ldots, w_n)$。
  - 从机器学习的角度来看：语言模型是对语句的概率分布的建模；
  - 通俗理解：语言模型判断一个语言序列是否是正常语句，如：

$$P(I\ am\ Light) > P(Light\ I\ am)$$

- 拼音转文本的本质被建模为一条隐含马尔科夫链，有很高的准确率。

第四部分

8.4 文本分类任务

8.3 文本分类任务

- 文本分类任务中一个分支任务：中文情感分析。
  - 在大众点评的评论中，希望得到关于评论情感倾向的分析，是“消极”或是“积极”；
  - 比如电影反馈时积极或消极的；
  - 在商品销售网站中的评价，根据1–5分辨别消费者态度……
- 态度使对特定的人或事物的带有主观色彩的偏好或倾向，情感分析是对态度的检测：
  - 持有人(来源)的态度；
  - 目标(方面)的态度；
  - 态度的类型：喜欢、讨厌……或者加入对态度的加权：很、非常……

8.3 文本分类任务

- 情感分析(Sentiment Analysis)可以分为：
  - 简单任务：这篇文章的态度是积极的还是消极的？
  - 更为复杂的：将文本中的态度按1-5的级别进行排序；
  - 高级的：检测目标、来源，或更复杂的态度类型。
- 简单将中文情感分析考虑为二分类问题，则完成任务步骤大致可分为：
  - 导入数据
  - 数据预处理
  - 模型训练
  - 模型测试

8.3 文本分类任务

- 导入数据：中文数据首先需要经过分词，在词中间加入空格：

comment	sentiment	cut comment
口味：不知道是我口高了，还是这家真不怎么样。？？我感觉口味确实一般很一般。	0	口味： 不 知道 是 我 口 高了， 还是 这家 真 不怎么样 。 ？ ？ 我 感觉 口味...
菜品丰富质量好，服务也不错！很喜欢！	1	菜品 丰富 质量 好， 服务 也 不错！ 很 喜欢！
说真的，不晓得有人排队的理由，香精香精香精香精，拜拜！	0	说真的， 不 晓得 有人 排队 的 理由， 香精 香精 香精 香精， 拜拜！
菜量实惠，上菜还算比较快，疙瘩汤喝出了秋日的暖意，烧茄子吃出了大阪烧的味道，想吃土豆片也是...	1	菜量 实惠， 上菜 还 算 比较 快， 疙瘩汤 喝出 了 秋日 的 暖意， 烧茄子 ...
先说我算是娜娜家风荷园开业就一直在这里吃？？每次出去回来总想吃一回？？有时觉得外面的西式简餐...	1	先说 我 算是 娜娜 家风 荷园 开业 就 一直 在 这里 吃 ？ ？ 每次 出去 回来 总 ...

8.3 文本分类任务

- 数据预处理，获取词向量(word2vec/word embeddings)，即将自然语言中的字词转为计算机可以理解的稠密向量。需要将中文经过以下预处理操作得到：
  - 将文档中出现过于频繁的词去掉；
  - 将文档中罕见词去掉；
  - 使用正则化处理掉文档中标点符号和数字等；
  - 设置停用词表（如虚拟词、冠词等）；
  - 使用word2vec模型得到文档的词向量。

8.3 文本分类任务

- 训练模型：可以使用机器学习或者深度学习中的方法，得到训练好的模型。
- 常用机器学习中的二分类方法：
  - 逻辑回归；
  - 朴素贝叶斯；
  - 支持向量机；
  - K近邻；
  - 支持向量机；
  - 随机森林；
- 深度学习中的方法：CNN、LSTM模型等。

8.3 文本分类任务

- 可能出现的问题：
  - 样本不平衡问题：不同极性的评论数量差距太大（例如，10^6好评和10^4差评），会导致分类器模型参数异常。解决方法为重抽样，使好评数与差评数均衡；或者采用代价敏感学习(cost-sensitive learning)，比如在训练SVM分类器时，将稀有样本错误分类的惩罚加大。
  - 处理打分问题（如1-5星）：可以将其转化为二元分类问题，小于2.5星视为负面评价，大于2.5星视为正面评价。或则和直接将星与极性强度用线性回归或其他方式拟合。