提纲

南京大学 NANJING UNIVERSITY

- **计算机视觉**
  - 图像形成
  - 简单图像特征
  - 图像分类
  - 物体检测
  - 三维世界

引言

南京大学 NANJING UNIVERSITY

- 大多数使用视觉的智能体采用**被动传感**，不需要主动发出光就能看到景象
- **主动传感**涉及雷达或超声波等信号的发射，以及对反射进行感知
- **特征**是通过对图像进行计算而获得的一串数字，用来表示图像
- 计算机视觉的两个核心问题
  - **重建**，智能体从一幅或一组图像中建立一个关于世界的模型，
  - **识别**，智能体根据视觉信息和其他信息对所见到的物体进行辨识

图像形成

南京大学 NANJING UNIVERSITY

无透镜成像：针孔照相机

- 图像传感器收集场景中物体散射的光，并构建二维图像
- 我们将整个图像平面称为一个传感器，但同时每个像素都是一个单独的微小传感器
- 聚焦图像：确保到达传感器的所有光子都来自世界中某个物 体上大致相同位置的点

针孔照相机:

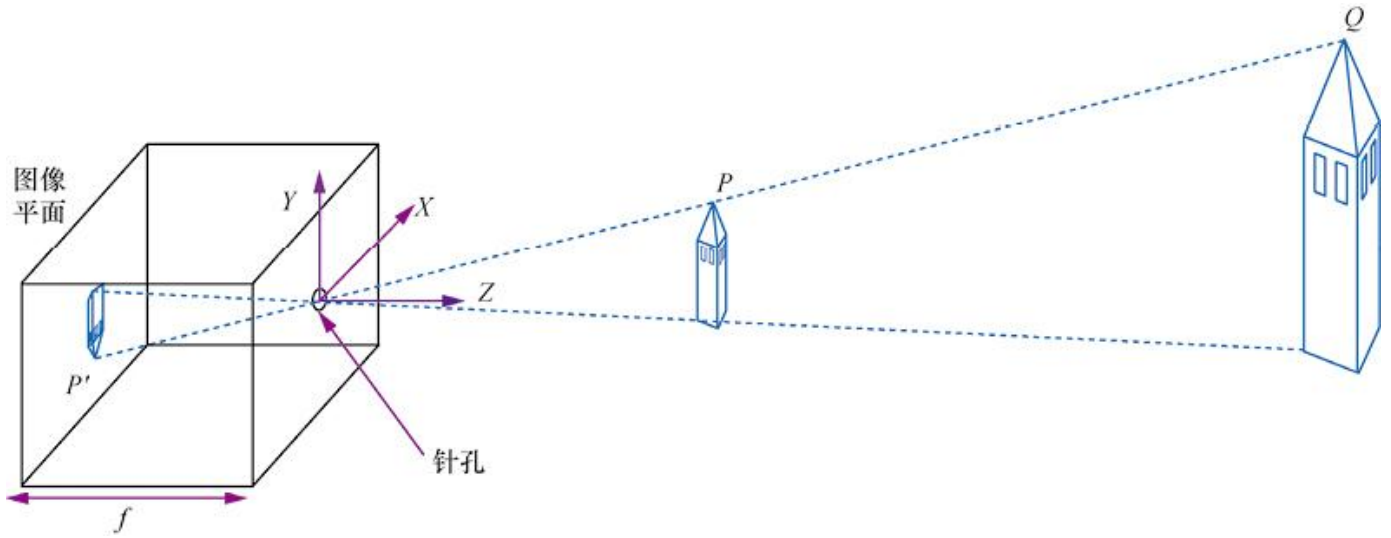
- 针孔照相机由盒子前部的针孔开口*O*和盒子后部的图像平面组成
- 这个开口也称作光圈，从针孔到图像平面的距离被称为**焦距**
- 只要物体在传感器的时间窗口内只移动一小段距离，我们也可以用针孔照相机来获得运动物体的聚焦图像
- 运动物体的图像会散焦，这种效应也称作**运动模糊**

图像形成

南京大学 NANJING UNIVERSITY

透视投影（perspective projection）

- 针孔照相机示意图



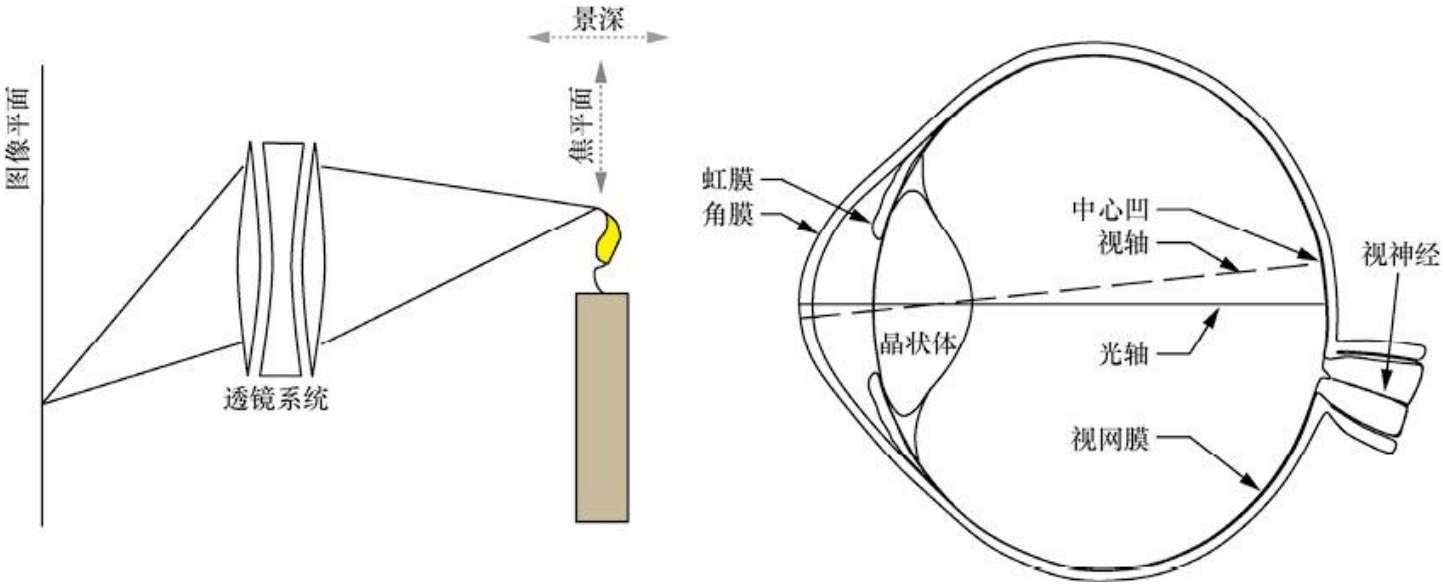
- 考虑场景中的点P，其坐标为(X, Y, Z)。P投影到图像平面上的点P′，其坐标为(x, y, z)。 若令f为焦距，即从针孔到像平面的距离，那么由相似三角形的性质可得

$$\frac{-x}{f}=\frac{X}{Z},\frac{-y}{f}=\frac{Y}{Z} \quad \Rightarrow \quad x=\frac{-fX}{Z},y=\frac{-fY}{Z}$$

图像形成

透镜系统

- 带透镜的照相机（或眼睛）将捕捉到所有照射到透镜上任何地方的光并将所有光聚焦到图像平面上的一个点
- 透镜的设计限制了它们只能聚焦距离透镜深度范围在Z以内的点上的光
- 这个范围的中心（聚焦最清晰的位 置）称作**焦平面**（focal plane）
- 聚焦能够保持足够清晰的深度范围称为**景深**（depth of field）
- 照相机的镜头光圈（开口）越大，景深越小



图像形成

南京大学 NANJING UNIVERSITY

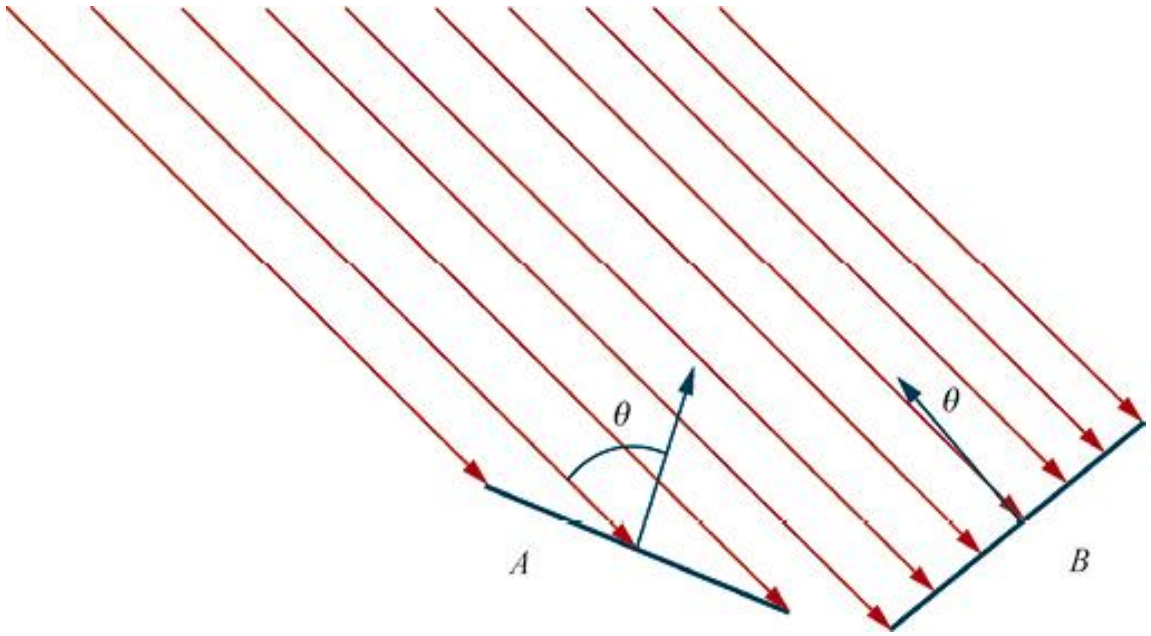
光线与明暗

- 图像中像素的亮度是关于投影到该像素的场景表面切片的亮度的函数
- 不明确性的产生是因为有3个因素影响了从物体上的一个点到达图像的光量:
  - 环境光（ambient light）的总光强
  - 该点处于向光面还是阴影中
  - 从该点反射（reflect）的光的总量
- **漫反射**将光均匀地散射到离开表面的各个方向上，因此漫反射表面的亮度不依赖于观察的方向
- **镜面反射**使得入射光以一定角度离开该表面，其方向由光到达的方向决定。
- 太阳是外界照明的主要来源，它的所有光线都从一个已知的方向平行地传播过来。
  - 采用**远点光源**来对这种行为进行建模

图像形成

南京大学 NANJING UNIVERSITY

光线与明暗



两个由远点光源照亮的表面切片，其中点光源的光线由带箭头的射线表示。

光线与明暗

- 由远点光源模型进行照明的漫反射表面切片将反射它收集到的光的一部分，具体的比例由**漫反射系数**给出。其范围通常是0.05 ～0.95。**兰伯特余弦定律**表明，漫反射切片的亮度由下式给出：

$$I = \rho I_0 \cos\theta$$

I_0为光源的光强，θ是光源方向和表面法线之间的夹角，ρ为漫反射系数。

- 各种照明的效果

图像形成

颜色

- 三原色原理
  - 通过混合适量的3个原色（primary），来达到任何光谱能量密度下的视觉效果，无论它有多么复杂。

- 三原色
  - 任何两种颜色的任意混合色都不会与第三种颜色相同.

- 常见的选择是红原色、绿原色以及蓝 原色，简称RGB

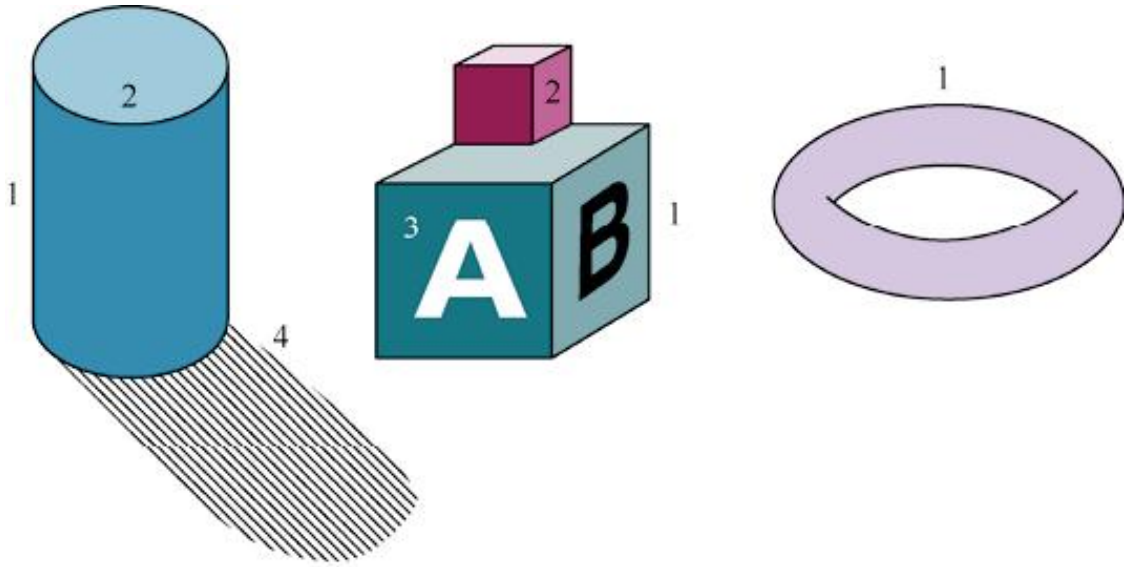
- 大多数计算机视觉应用将一个表面建模为具有3种不同（RGB）漫反射系数，并将光源建模为具有3种（RGB）强度的模型。在此基础上，将兰伯特余弦定律应用于每个像素，以获得红、绿和蓝的像素值。

简单图像特征

南京大学 NANJING UNIVERSITY

边缘

- 边缘是图像平面中的直线或曲线，图像亮度在此处发生了“显著”的变化。



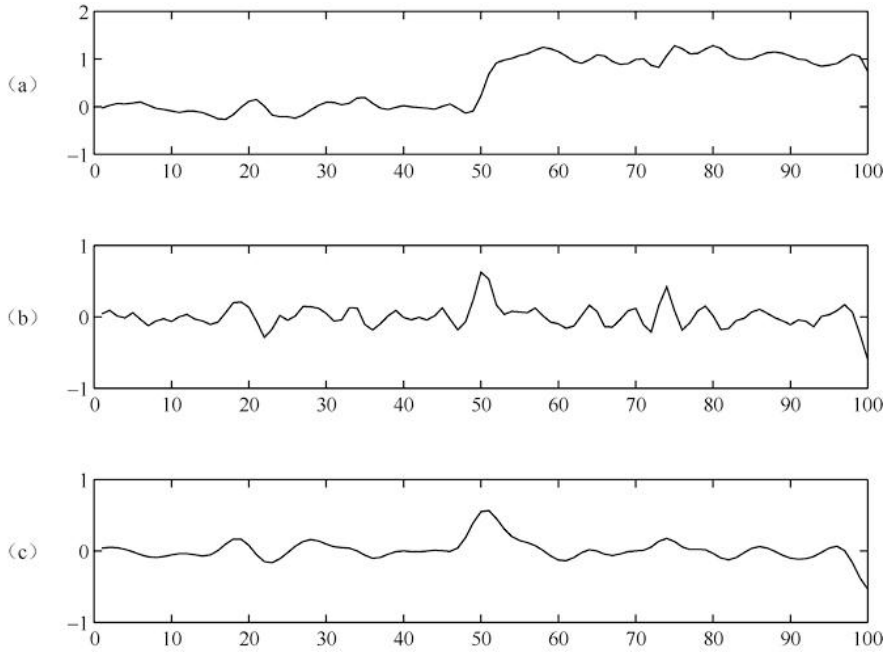
不同类型的边缘：（1）深度不连续；（2）表面方向不连续；（3）反射不连续；（4）照明不连续（阴影）

简单图像特征

南京大学 NANJING UNIVERSITY

边缘

- 对图像进行微分，然后寻找导数I'(x)较大的位置来对图像边缘进行辨别
  - 仅根据光强差异会因噪声而导致识别边缘时出现错误
  - **噪声**：像素值的变化与边缘无关



（a）一个跨过边缘的一维截面上的光强$I(x)$。（b）强度的导数$I'(x)$。此函数中的较大值对应于边缘，但函数中存在噪声。（c）经过平滑处理后的光强的导数。

简单图像特征

南京大学 NANJING UNIVERSITY

边缘

- 使用周围像素来控制噪声，从而进行平滑处理
- 使用附近像素的加权和作为 对像素“真实”值的预测，其中距离最近的像素的权重最大
- **高斯滤波器：**

$$G_\sigma(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/(2\sigma^2)} \quad \text{（一维情况）}$$

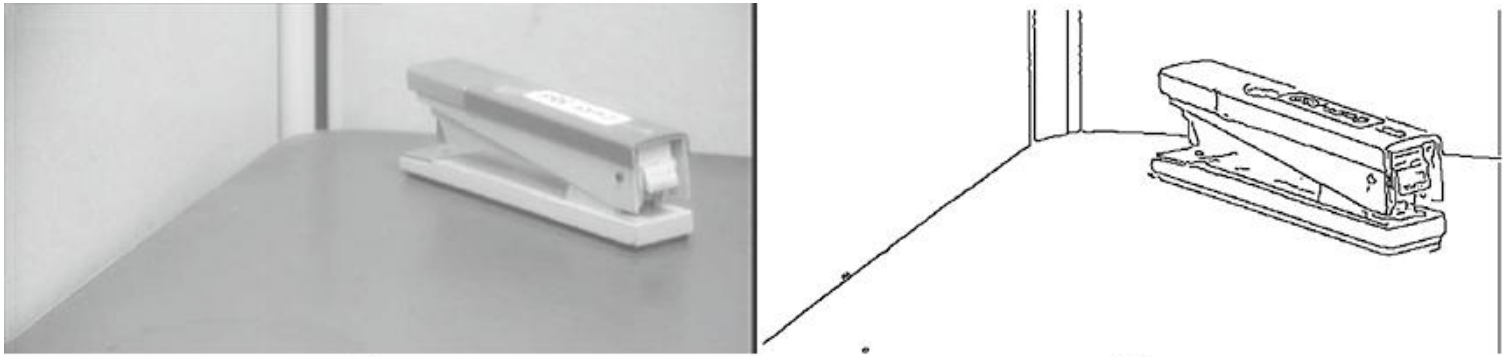
$$G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/(2\sigma^2)} \quad \text{（二维情况）}$$

- 应用高斯滤波器意味着将光强$I(x_0, y_0)$替换为 $I(x, y)G_\sigma(d)$ 在所有像素 (x, y) 上的和, 其中 d 表示从 (x_0, y_0) 到 (x, y) 的距离。
- 平滑函数可以表示为图像与高斯核的卷积，进而可以将平滑和边缘检测融合到一个操作中，因为卷积的导数等于两函数之一的导数与另一函数进行卷积。

简单图像特征

南京大学 NANJING UNIVERSITY

边缘



（a）一个订书机的照片。（b）从（a）中计算出的边缘

简单图像特征

南京大学 NANJING UNIVERSITY

纹理

- 在计算视觉中，**纹理**是指表面上可以被视觉感知到的图案。通常来说，这些图案大致上是规则的。
- 常见的粗糙的纹理模型是元素的模式重复，有时也称为**纹理元素**或**纹元**
- 纹理是图像切片的性质，而不是一个孤立像素的特性
- 纹理表示已经在两个重要任务中发挥了较大作用：物体识别和图像块匹配
- 纹理表示的一个基本构造方法
  - 给定一个图像切片，计算该切片中每个像素的梯度方向
  - 然后利用关于方向的直方图对该切片进行表征

简单图像特征

南京大学 NANJING UNIVERSITY

光流

- **光流:** 当照相机与场景中的一个或多个物体之间发生相对运动时，图像中产生的视运动。它描述了观察者和场景之间的相对运动而导致的图像中特征的运动方向和速度。

- 相似性度量：**差值平方和(*SSD*)**

$$\mathrm{SSD}(D_x, D_y) = \sum_{(x,y)} (I(x,y,t) - I(x+D_x, y+D_y, t+D_t))^2 .$$

- 其中, (x, y) 表示以为(x_0, y_0)中心的像素块中的像素位置。寻找 (D_x, D_y)使SSD最小化。 (x_0, y_0)处的光流 $(v_x, v_y) = (D_x/D_t, D_y/D_t)$.

- 场景中应该存在一些纹理，从而使得图像中像素之间存在亮度的显著差异。

简单图像特征

南京大学 NANJING UNIVERSITY

光流

一个视频序列的两帧，以及与从一帧到另一帧的位移相对应的光流场。注意由箭头方向刻画的网球拍和右腿的动作。

简单图像特征

南京大学 NANJING UNIVERSITY

自然图像分割

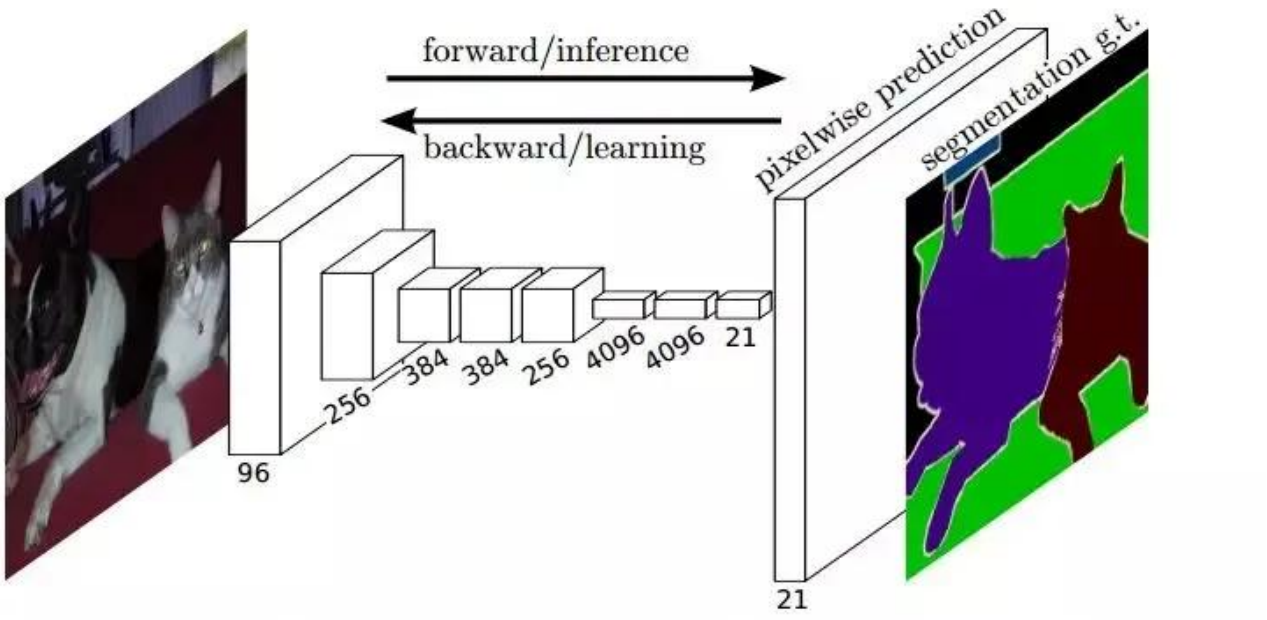
- **分割**是将一幅图像分解成若干组相似像素集的过程

- 两种方法: 检测边界或将像素聚类成多个区域
  - 分类问题
  - 像素聚类问题

（a）原始图像。（b）图像的边界轮廓。（c）通过对图像精细划分得到的各个分割区域。（d）通过对图像进行较粗糙的分割得到的各个分割区域。

简单图像特征

自然图像分割



Fully Convolutional Networks for Semantic Segmentation (CVPR 2015)

图像分类

- 现代的计算机视觉系统通过外观（例如颜色和纹理，而不通过几何性质）对图像进行分类。

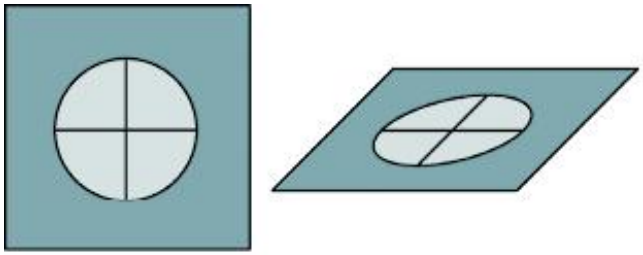
- 存在两个难点：

- 同一类别的不同实例可能看起来不同，也就是存在类内差异

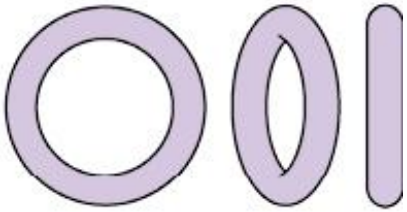
- 不同的时刻看起来可能不同，取决于以下几种效应
  - 光照
  - 透视收缩
  - 视角
  - 遮挡
  - 变形

- 现代方法通过使用卷积神经网络从大量的训练数据中学习表示和分类器来处理这些问题。在训练集足够丰富的情况下，分类器在训练中会多次看到任何一个重要的效应， 因此可以根据具体效应进行调整。

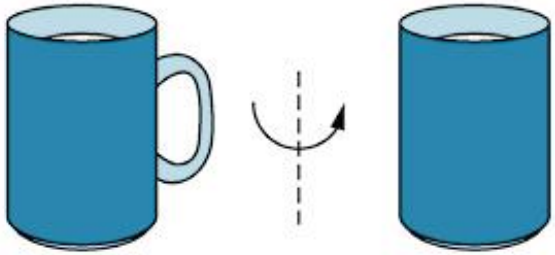
图像分类



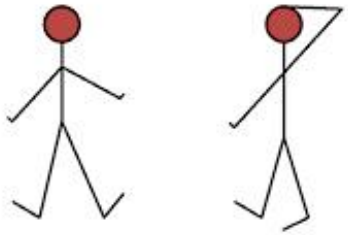
透视收缩



视向



遮挡



变形

产生外观变化的重要因素，它们可使同一物体的不同图像看起来不同。

图像分类

南京大学 NANJING UNIVERSITY

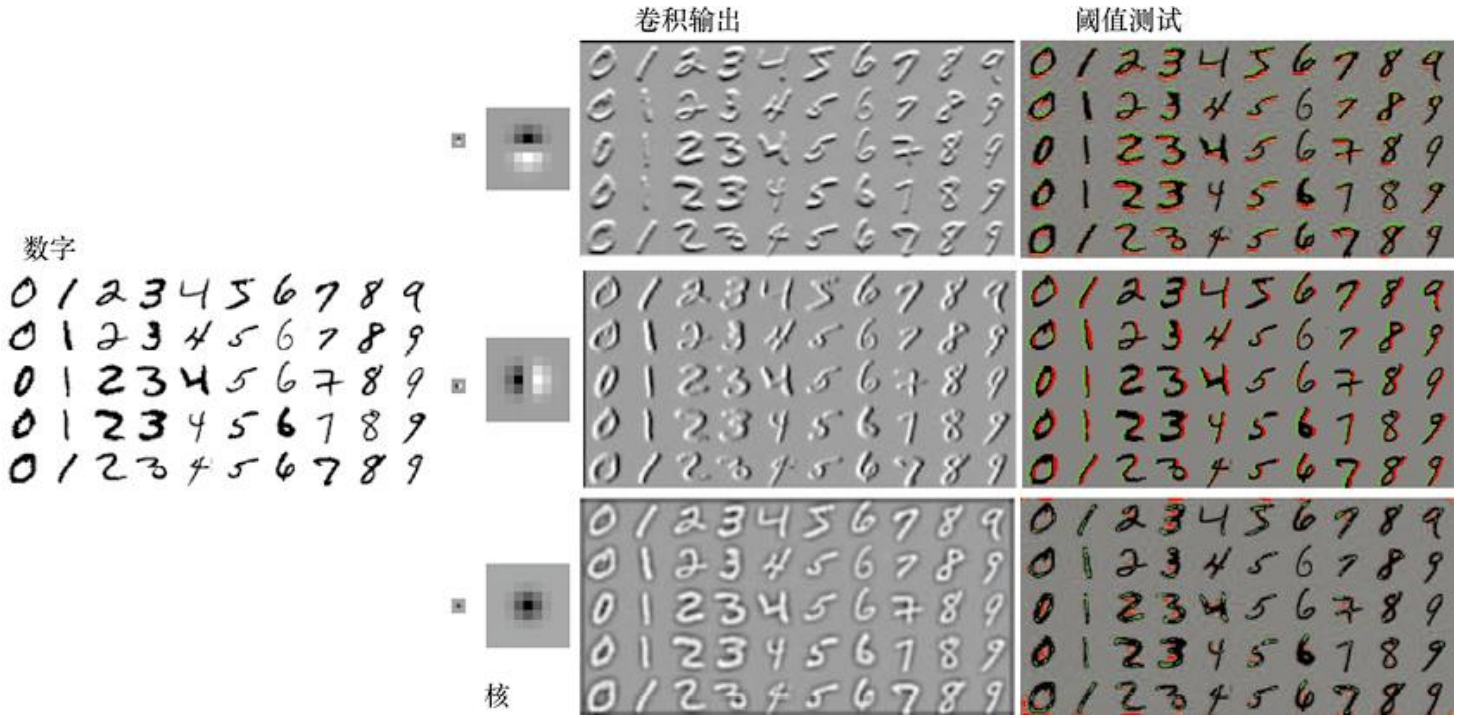
基于卷积神经网络的图像分类

- 卷积神经网络（CNN）是非常成功的图像分类器。在有足够的训练数据和较好的训练技巧情况下，CNN产生了非常成功的分类系统
- 可以拍摄一幅关于数字的图像，并在不改变数字本身的情况下进行一些小的更改
- 局部模式可以提供相当多的信息
- 局部模式之间的空间关系也包含较多信息
- 卷积与ReLU 激活函数的复合看作一个局部模式检测器
  - 卷积将度量图像的每个局部窗口与核模式的相似程度
  - ReLU激活函数将低分窗口置为零，并突出高分窗口
- 多个卷积核的卷积可以找到多种模式
- 可以通过将新的一层应用于第一层的输出来检测复合模式

图像分类

南京大学 NANJING UNIVERSITY

基于卷积神经网络的图像分类



最左侧是MNIST数据集中的一些图像。中间图的左侧为3个卷积核。它们以实际大小（图中的小方块）给出，并放大以显示其内容：中度灰色的值为0，浅色表示正值，深色表示负值。中间图的右中侧给出了将左侧这些核应用于图像的结果。最右侧给出了响应大于阈值（绿色）与小于阈值（红色）的像素。

图像分类

基于卷积神经网络的图像分类

- **数据集增强：**对训练样本进行复制并稍加修改
- 将图像随机地移动、旋转或稍微拉伸， 或者将像素的色彩随机地进行少量调整
- 也可以把数据集增强的方法用于测试过程，而不是训练过程。
- 基于CNN 的分类器擅长忽略那些没有区分力的模式
- **环境或者上下文（context）：**物体上的模式可能是有区分力的
  - 例如一个猫玩具、一个带小铃铛的项圈或者一盘猫粮实际上可能有助于我们判断出所观察的物体是猫

物体检测

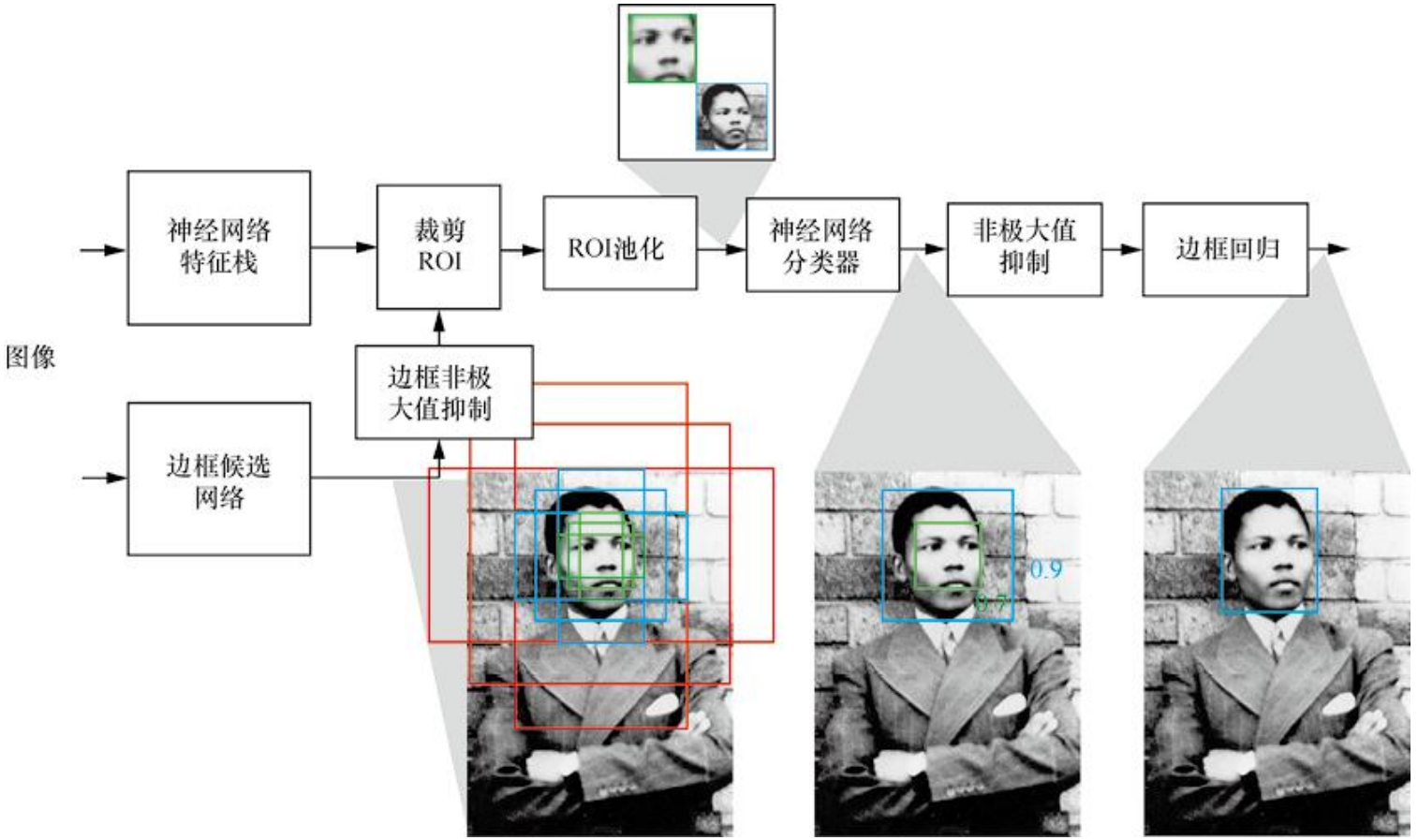
- **物体检测**器在一幅图像中寻找多个物体，分别判断每个物体属于什么类别，并通过在物体周围添加一个边框来反映出每个物体的位置。

- 构建一个物体检测器:
  - 可以通过在较大的图像上观察一个小的滑动窗口（一个矩形）
  - 在每个检测点上，我们使用CNN分类器对窗口中观测到的内容进行分类

- 细节:
  - 确定窗口的形状
  - 为窗口构建一个分类器
  - 决定要查看哪些窗口
  - 选择要报告的窗口
  - 利用这些窗口反映物体的精确位置

- 一个能找到包含物体的区域的网络称为**区域候选网络**（regional proposal network，RPN）
  - *Faster RCNN* 将大量边界框集合编码为固定大小的映射

物体检测



Faster RCNN使用两个网络： 一个网络用于计算候选图像框（称为“锚框”）的物体检测得分； 第二个网络是一个特征栈，用于计算适合分类的图像表示。

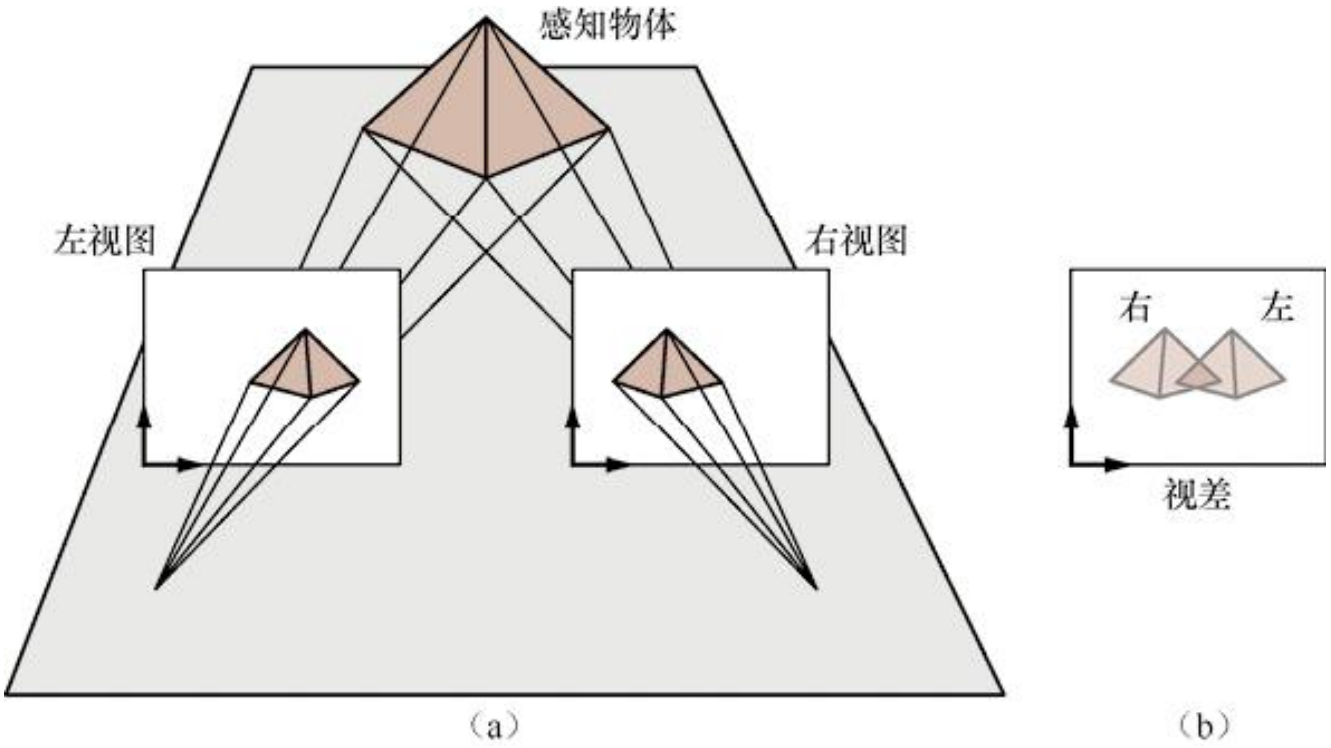
三维世界

- 在三维世界中，有两张关于同一物体的图片通常比只有一张要好：
- 如果你从不同的视角拍摄了同一场景的两幅图像，并且你对这两部摄像机了解得足够多，那么你可以通过计算第一个视图中的点对应第二个视图中的哪个点，并应用一些几何知识，来构建一个三维模型
- 如果你有两个包含足够多点的视图，并且你知道第一个视图中的点对应第二个视 图中的哪个点，那么你不需要对摄像机了解太多就可以构建出该三维模型
- 关键的问题在于建立第一个视图中的点与第二个视图中的点的对应关系
- 通常有两种方法来获得一个场景的多个视图
  - 安置两部摄像机
  - 移动摄像机

三维世界

南京大学 NANJING UNIVERSITY

双目立体视觉



将摄像机按平行于图像平面的方式进行平移会导致图像特征在摄像机平面中发生移动。（a）位置上的差异是对物体深度的暗示。（b）如果我们对左右两幅图像进行叠加，我们将观察到视差。

三维世界

单个视图的三维线索

- 如果图片中有证据表明一个物体遮挡了另一个物体，那么遮挡另一个物体的物体将离眼睛更近。
- 纹理也是三维结构的重要线索。尽管纹理元素在场景中的物体上的分布可能是均匀的，但在透视图中，由于透视收缩，纹理看起来将是不均匀的。
- 明暗是三维形状的一个线索，如果一个表面的法线指向光源，则该表面会更亮；如果它背向光源，那么该表面会较暗。
- 物体之间的空间关系是另一个重要线索。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

理解人类行为

从单一的图像中重建人类模型。**最左图**为一张图片，**中左图**为原图与重建出的身体叠加的图片，**中右图**为重建出的身体的另一个视图，**最右图**是重建出的身体的另一个不同视图。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

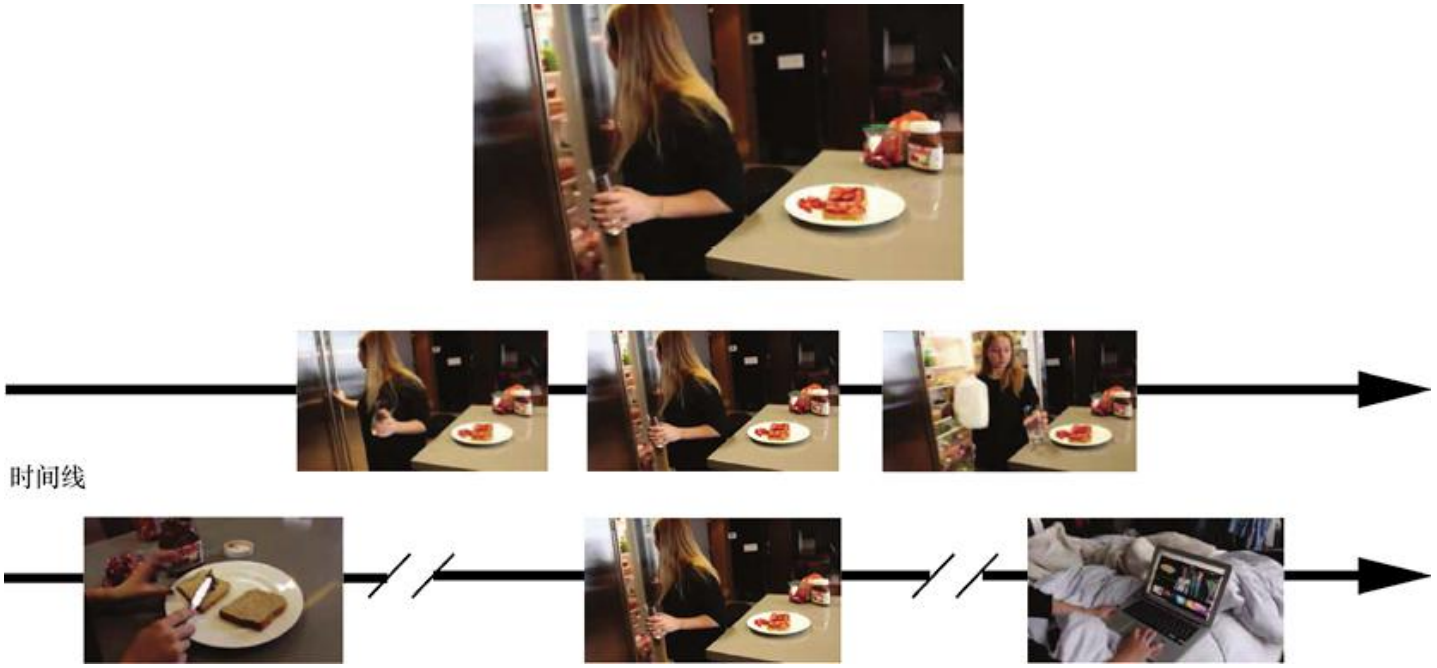
理解人类行为

同一个动作看起来很不一样，不同的动作看起来很相似。这些例子是来自一个数据集中的自然动作。上面3幅图表示标签为“打开冰箱”的样本，有的是特写，有的是远处拍摄。下面3幅图表示标签为“从冰箱里拿东西”的样本。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

理解人类行为



我们所说的动作取决于时间尺度。对于最上面的单幅图像，最好的描述是“打开冰箱”。但是，如果你看完了一段视频短片（由中间一行图像表示），关于这个动作的最佳描述就是“从冰箱里拿牛奶”。如果你看完了一段较长的视频（由最下面一行图像表示），关于这个动作的最佳描述是“准备点心”。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

匹配图片与文字

A baby eating a piece of food in his mouth

A young boy eating a piece of cake

A small bird is perched on a branch

A small brown bear is sitting in the grass

自动图像标题系统给出了一些好的结果和一些失败的结果。左边的两个标题很好地描述了各 自的图像，尽管“eating … in his mouth”是一个不流畅的表达，这是早期标题系统所使用的循环神经网 络语言模型的一个相当典型的特点。根据右边的两个标题，我们认为标题系统似乎不了解松鼠，所以 从环境猜测该动物；它也没有意识到这两只松鼠在吃东西。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

匹配图片与文字

视觉问答系统产生关于图像的自然语言问题的答案。**顶部：**该系统对有关图像的一些相当棘手的问题给出了非常合适的答案。**底部：**不太令人满意的答案。例如，系统被要求猜测比萨饼上的洞的个数，但系统并不知道什么算洞，而且洞本身很难计数。类似地，系统认为猫的腿的颜色是棕色，这是因为图片背景是棕色的，并且系统不能正确定位猫的腿。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

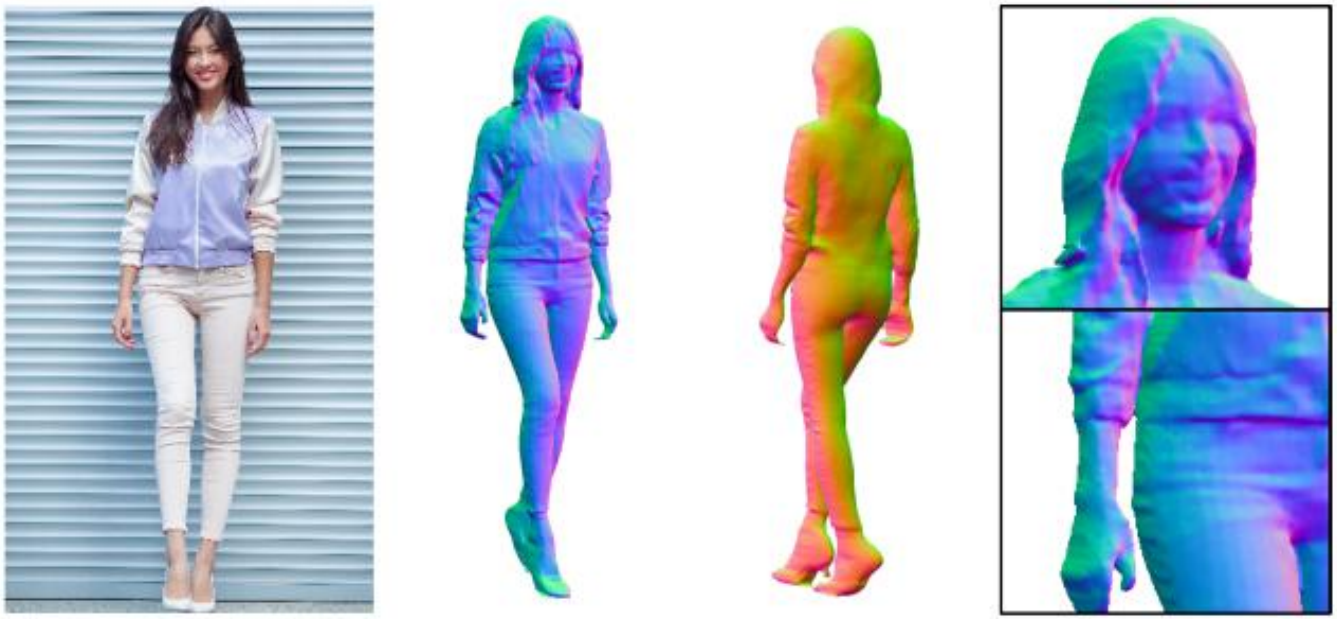
多视图重建

Representing Scenes as Neural Radiance Fields for View Synthesis
(ECCV 2020)

计算机视觉的应用

南京大学 NANJING UNIVERSITY

单视图中的几何

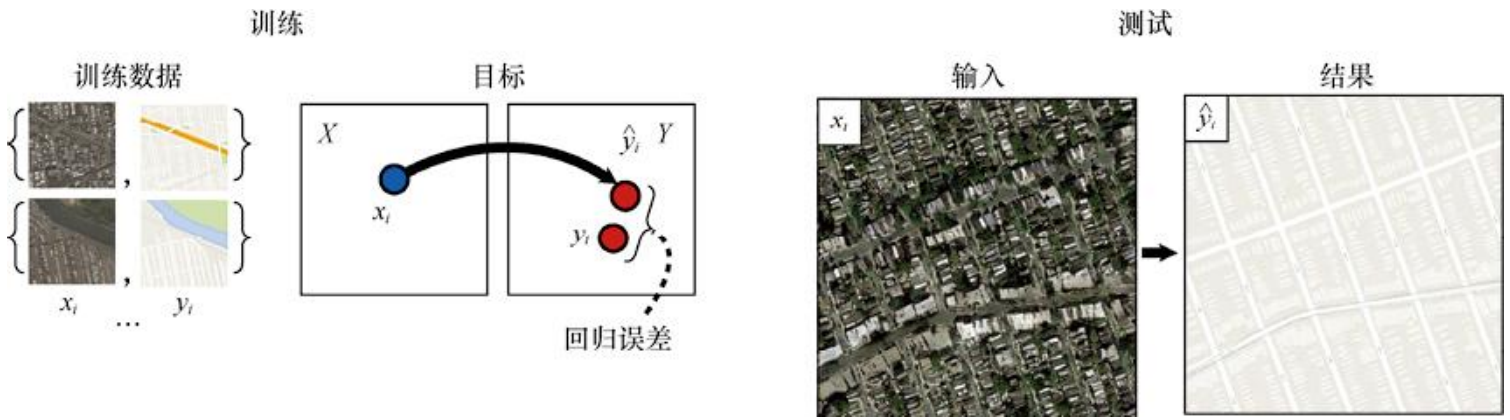


PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization (CVPR 2020)

计算机视觉的应用

南京大学 NANJING UNIVERSITY

生成图片

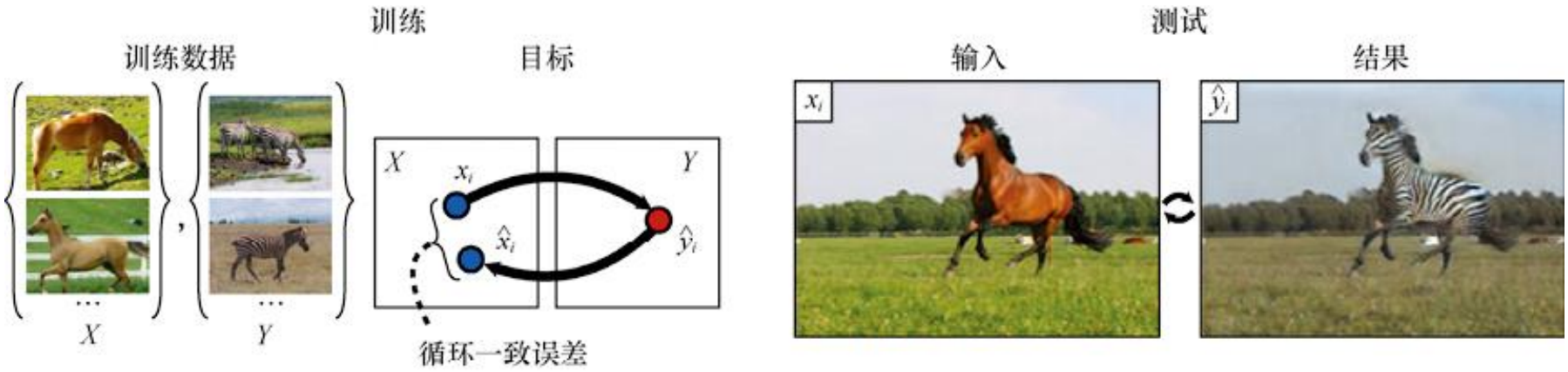


成对图像的转换，其中输入由航空影像和相应的道路图组成，我们的目标是训练一个从航空影像生成道路图的网络（该系统还可以学习从道路图生成航空影像。）网络通过比较 （X型样本xi的输出）和Y型的正确输出yi进行训练。在测试时，网络必须从新的X型输入中生成新的Y型图像。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

生成图片



未配对图像转换：给定两组图像（X型是马，Y型是斑马），但没有对应的配对，我们要学习将马转换成斑马。该方法训练两个预测器：一个将X型映射为Y型，另一个将Y型映射为X型。如果第一个网络将马xi映射为斑马，那么第二个网络应当把 映射回原始的xi。两个网络利用xi和 i之间的差进行训练。从Y型到X型再回到Y型的循环必须是封闭的。这样的网络可以成功地对图像进行丰富的变换。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

生成图片

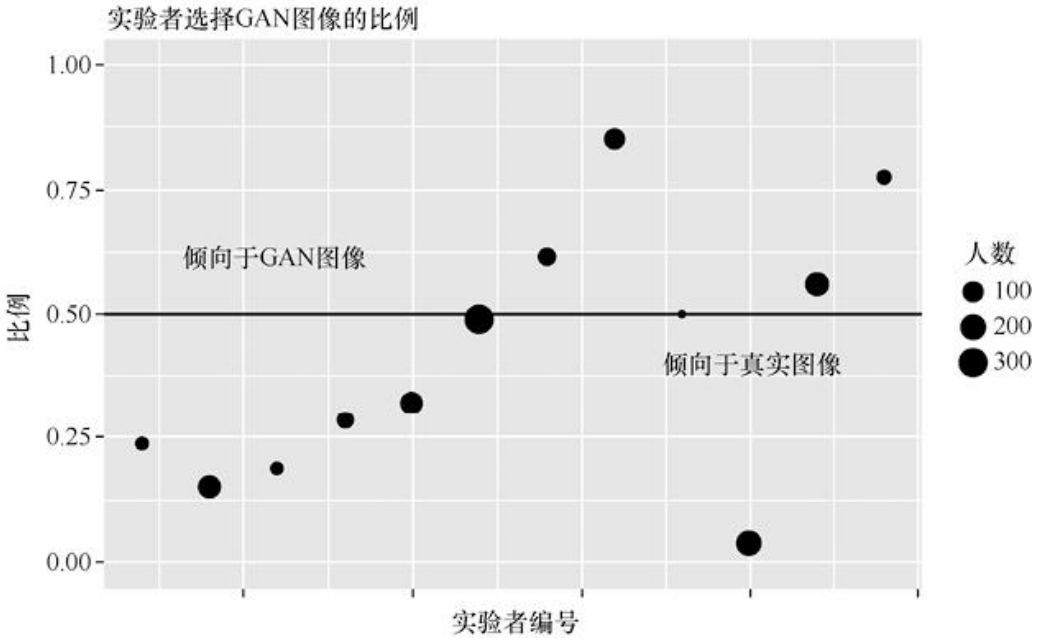


风格转换：将内容为猫的照片与抽象绘画的风格相结合，生成经过抽象风格渲染的猫的新图像（右图）。

计算机视觉的应用

南京大学 NANJING UNIVERSITY

生成图片



GAN生成了肺部X射线图像。左图的一对图像为一张真实的X射线图和一张由GAN产生的X射线图。右图为一项测试的结果，它要求放射科医生在看到左边所示的一对X射线图后，判断哪一张X射线图是真实的。平均来说，他们的选择正确率为61%，这比任意猜测的结果要好一些。

小结

- 图像的表示蕴含着边缘、纹理、光流和区域等信息。这些信息为我们提供了有关物体边界以及图像之间对应关系的线索。
- 利用卷积神经网络可以得到精确的图像分类器，它从大量数据中去学习对分类任务有用的特征。
- 图像分类器可以转换成物体检测器。一个分类器对图像中框内的内容进行评分， 另一个分类器则判断该框中是否包含物体，以及该物体是什么。
- 当我们有同一个场景的多个视图时，我们可以重建场景的三维结构以及视图之间 的关系。在许多情况下，我们也可以从单个视图中重建三维几何结构。
- 计算机视觉方法的应用非常广泛。