



# 人工智能基础与进阶

## 深度网络实践

上海交通大学

# 目录 content



## 第一节 深度学习基础模型

## 第二节 深度学习案例介绍

## 第三节 深度学习案例演示

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

# 第一节 深度学习基础模型

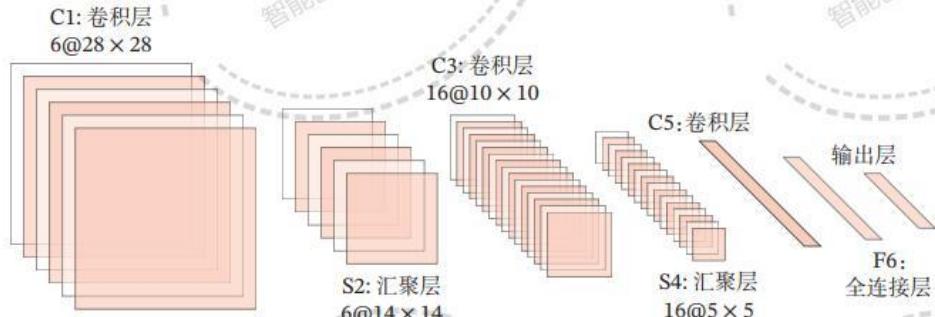
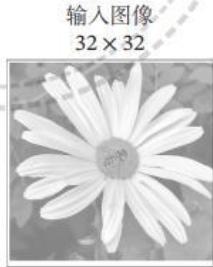


上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

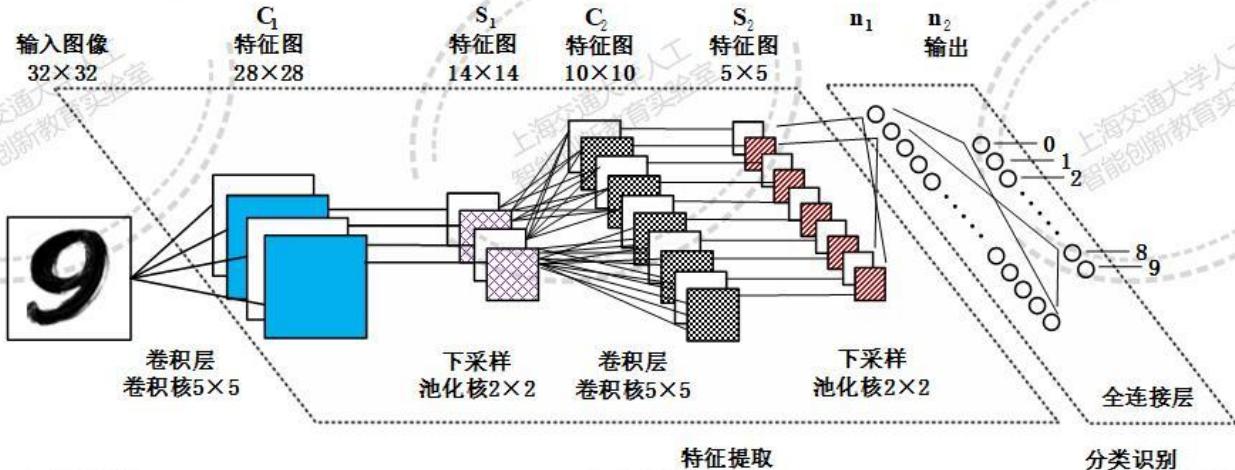
上海交通大学人工  
智能创新教育实验室

# LeNet-5



LeNet-5是一个非常成功的神经网络模型，基于LeNet-5的手写数字识别系统在20世纪90年代被美国很多银行使用，用来识别支票上面的手写数字。

# LeNet-5



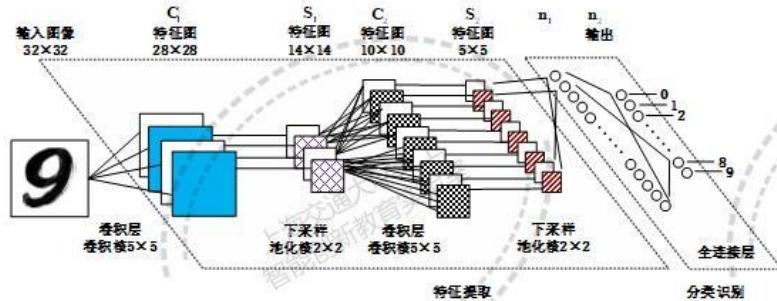
**输入：**  $32 \times 32 = 1024$  的手写字体。这些手写字体包含 0~9 数字，也就是相当于 10 个类别的图片

**C1 层：** 作者选择了 6 个特征卷积核，然后卷积核大小选择  $5 \times 5$ ，这样我们可以得到 6 个特征图，然后每个特征图的大小为  $32 - 5 + 1 = 28$

**S2 层：** 下采样层，也就是使用最大池化进行下采样，池化的滤波器大小选择  $(2, 2)$ ，步长 stride 为 2。这样我们可以得到输出大小为  $14 \times 14$ ，且有 6 个这样的图片

**C3 层：** 卷积层，这一层我们选择卷积核的大小依旧为  $5 \times 5$ ，据此我们可以得到新的图片大小为  $14 - 5 + 1 = 10$ ，此处采用 16 个卷积核，所以最终输出 16 个  $10 \times 10$  的图片

# LeNet-5



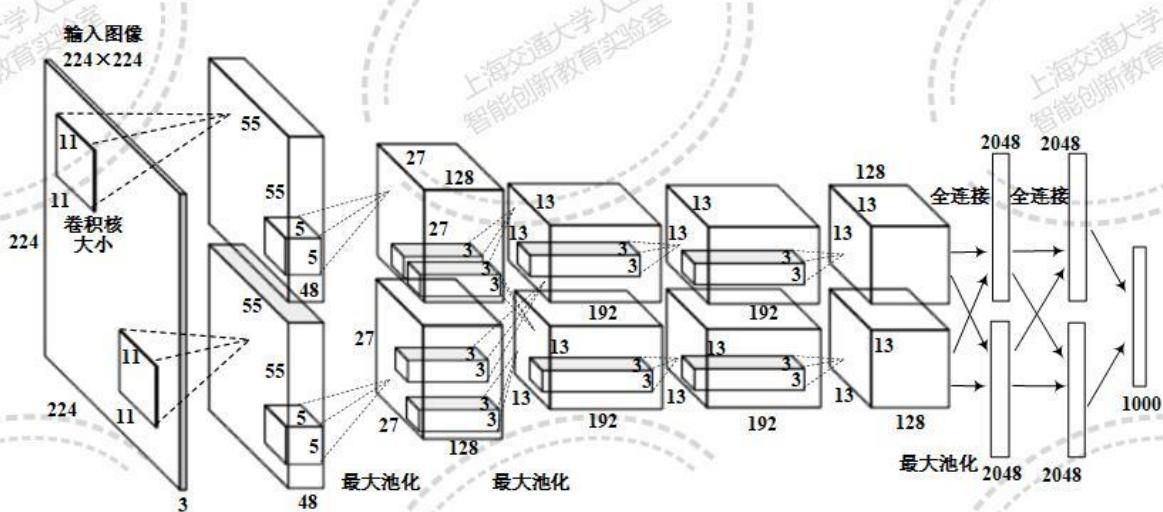
**S4层：**下采样层，对C3的 $16 \times 10 \times 10$ 的图片进行最大池化，池化的滤波器大小选择 $(2, 2)$ ，步长stride为2。因此最后S4层为16张大小为 $5 \times 5$ 的图片

**C5层：**将S4层的输出平铺为一个400的一维向量。然后用这400个神经元构建下一层，C5层有120个神经元。S4层的400个神经元与C5层的每一个神经元相连【C5层有120个神经元】，这就是全连接层，可看作一个标准的神经网络层

**F6层：**对C5层的120个神经元再添加一个全连接层【F6层含有84个神经元】

**输出：**最后将F6层的84个神经元填充到一个SoftMax函数，得到输出长度为10的张量，张量中为1的位置代表所属类别。（例如[0,0,0,1,0,0,0,0,0]的张量，1在index=3的位置，故该张量代表的图片属于第三类）

# AlexNet



AlexNet是第一个现代深度卷积网络模型，其首次使用了很多现代深度卷积网络的技术方法，比如使用两个GPU进行并行训练，采用了ReLU作为非线性激活函数，使用Dropout防止过拟合，使用数据增强来提高模型准确率等。AlexNet赢得了2012年ImageNet图像分类竞赛的冠军。

# AlexNet

AlexNet的输入为 $224 \times 224 \times 3$ 的图像(预处理后图像尺寸 $227 \times 227 \times 3$ )，输出为1000个类别的条件概率，具体结构如下：



## AlexNet中的trick

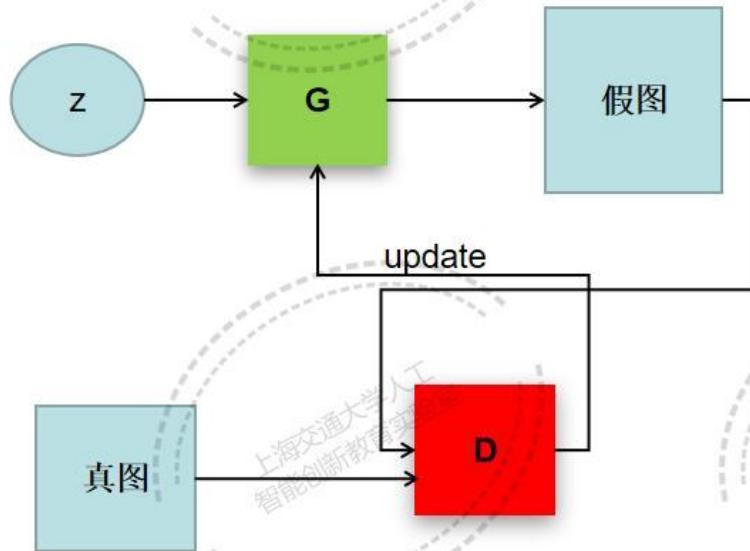
AlexNet将CNN用到了更深更宽的网络中，其效果分类的精度更高相比于以前的LeNet，其中有一些技巧是必须要知道的：

- **ReLU的应用：**AlexNet使用ReLU代替了Sigmoid，其能更快的训练，同时解决sigmoid在训练较深的网络中出现的梯度消失的问题
- **Dropout随机失活：**随机忽略一些神经元，以避免过拟合
- **提出了LRN层：**局部响应归一化，对局部神经元创建了竞争的机制，使得其中响应小的值变得更大
- **使用了GPU加速计算：**使用了gpu加速神经网络的训练
- ...

# 生成对抗网络 (GAN)

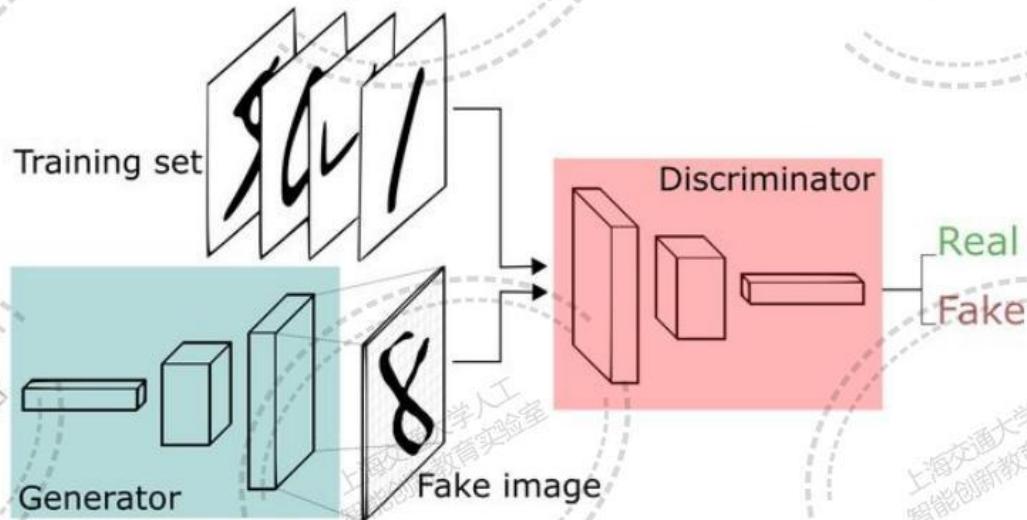
G是generator，生成器：负责凭空捏造数据出来

D是discriminator，判别器：负责判断数据是不是真数据



# 生成对抗网络

GAN的主要结构包括一个生成器Generator 和 一个判别器Discriminator



The background features a robotic arm with blue and silver components interacting with a human hand. The scene is set against a light blue gradient with faint, repeating circular patterns containing the text "上海交通大学人工智能创新教育实验室".

## 第二节 深度学习案例介绍

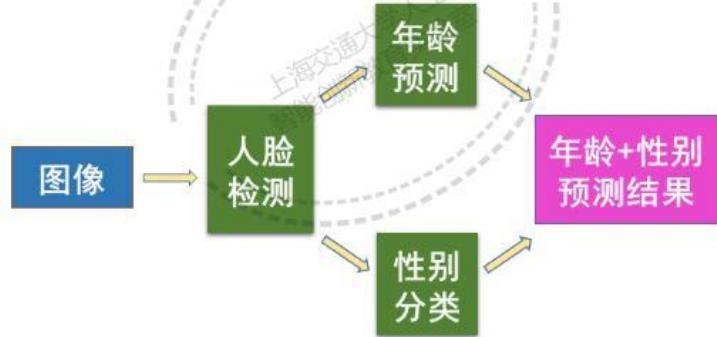
# 情绪检测



待测图像输入程序后，会经过人脸检测确定图像中的所有人脸，确定为感兴趣区域，之后经过卷积神经网络提取出图像中的特征，然后再利用特征对情绪进行分类，判断出各种情绪的置信度。最后将分类结果与人脸检测结果结合，将不同的情绪标签绘制在人脸附近，我们就能看到带着情绪标签的图像了。

对于视频而言则是需要对每一帧图像做这样的操作，我们就能看到情绪检测实时的演示效果。

# 年龄性别检测



待测图像输入程序后，会经过人脸检测确定图像中的所有人脸，确定为感兴趣区域并在图像中框出，然后再将人脸的图像及人脸特征输入到年龄预测及性别分类的模块，进行人脸的年龄性别预测。最后将预测结果的标签绘制在检测到的人脸附近，我们就能看到框出人脸并标记性别和年龄的图像了。

对于视频而言则是需要对每一帧图像做这样的操作，我们就能看到性别年龄检测实时演示效果。

# 风格迁移

风格



风格迁移:保持图像的内容不变，并且将图像转变为某种指定的风格

# 风格迁移



原始图像输入程序后，会经过预先训练的风格迁移网络，保持图像的内容不变，并且将图像转变为某种指定的风格。

对于视频而言则是需要对每一帧图像做这样的操作，我们就能看到风格迁移实时演示效果。

The background features a robotic arm with blue and silver components interacting with a human hand. The scene is set against a light blue gradient with faint, repeating circular patterns containing the text "上海交通大学人工智能创新教育实验室".

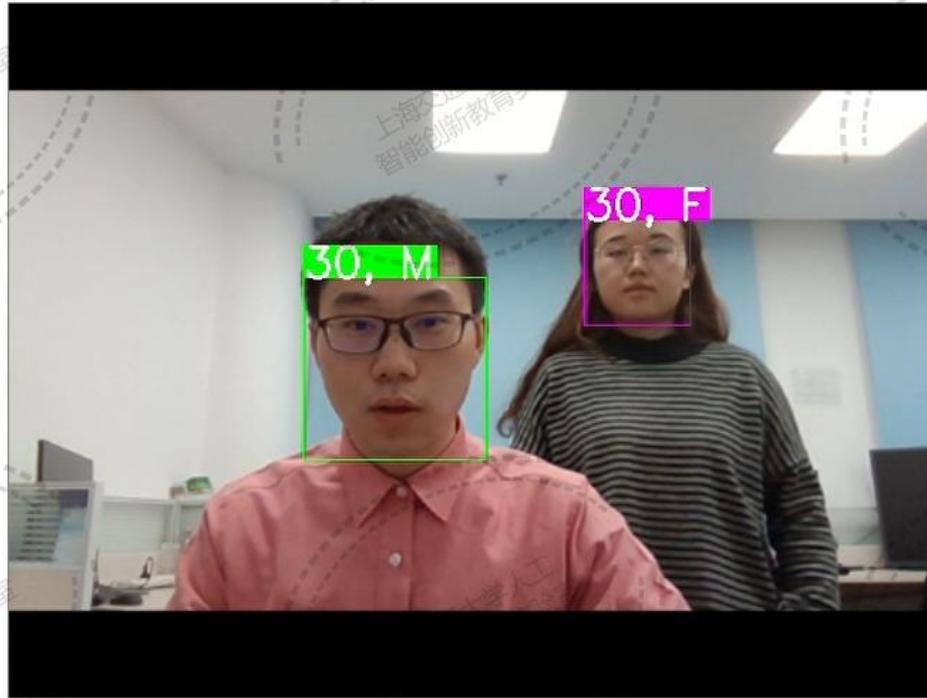
## 第三节 深度学习案例演示

# 情绪检测演示



情绪检测：根据人的面部的表情判断其情绪，并在图像中标记出来

# 年龄性别检测演示



年龄性别检测：在图像中，通过面部判断人的性别和年龄，并在图像中标出

# 风格迁移演示



风格迁移：将输入图像的风格转变为其他风格，并保持图像内容不变



谢 谢 聆 听

THANKS FOR YOUR ATTENTION