## 机器学习模型

1. 通用物体检测和识别（General Object Detection/Recognition）

传统方法：

1. 基于Bag Of Words词袋模型的，SIFT/SURF+KMeans+SVM
2. 基于Sparse Coding稀疏编码的，LLC
3. 基于聚合特征的，Fisher Vector/VLAD
4. 基于变形部件组合模型的，DPM用到HOG/Latent SVM

相关论文：

1、Visual Object Recognition, Kristen Grauman

2、Locality-constrained Linear Coding for Image Classification

3、Fisher Kernels on Visual Vocabularies for Image Categorization

4、Improving the Fisher Kernel for Large-Scale Image Classification

5、Aggregating local descriptors into a compact image representation

6、Object Detection with Discriminatively Trained Part Based Models

相关开源地址：

<http://www.vlfeat.org>

<https://github.com/rbgirshick/voc-dpm>

<https://github.com/cbod/cs766-llc>

深度学习：

RCNN/SPPNet/Faster RCNN，Yolo系列，SSD，R-FCN，RetinaNet，CFENet。

相关论文：

1、Rich feature hierarchies for accurate object detection and semantic segmentation

2、Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition

3、Fast R-CNN

4、Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

5、You Only Look Once: Unified, Real-Time Object Detection

6、YOLO9000: Better, Faster, Stronger

7、YOLOv3: An Incremental Improvemen

8、SSD: Single Shot MultiBox Detector

9、R-FCN: Object Detection via Region-based Fully Convolutional Networks

10、Focal Loss for Dense Object Detection

11、CFENet: An Accurate and Efficient Single-Shot Object Detector for Autonomous Driving

相关开源地址：

<https://github.com/rbgirshick/rcnn>

<https://github.com/rbgirshick/fast-rcnn>

<https://github.com/rbgirshick/py-faster-rcnn>

<https://github.com/balancap/SSD-Tensorflow>

<https://github.com/chuanqi305/MobileNet-SSD>

<https://github.com/gliese581gg/YOLO_tensorflow>

<https://github.com/choasup/caffe-yolo9000>

<https://github.com/qqwweee/keras-yolo3>

<https://github.com/daijifeng001/R-FCN>

<https://github.com/YuwenXiong/py-R-FCN>

<https://github.com/daijifeng001/caffe-rfcn>

<https://github.com/facebookresearch/Detectron>

1. 特定物体检测和识别和检索（Specific Object Detection/CBIR）

特定物体只识别一张特定的图，不能进行大样本训练，也即不需要进行训练和学习。大多数只是用Artificial Feature手工特征，比如特征点，而且对于刚性物体，特征点匹配可以用SVD分解计算出仿射变换矩阵，进而判断物体边缘的方向。

1. 特征点匹配，基于欧氏距离的，如SIFT/SURF，基于海明距离的，如AKAZE/FREAK，欧氏距离的检索可以用KD-Tree或者其他算法如hnsw、Falconn，海明距离的检索用LSH。
2. 基于Fisher Vector/VLAD，采用随机超平面的方式切换成海明距离进行检索
3. 基于神经网络的，sPoC特征，R-MAC，NetVlad用的都是预训练模型。
4. 检索，基于欧式距离的检索有hnsw、Falconn、Faiss等开源库。

相关论文：

Aggregating Deep Convolutional Features for Image Retrieval

PARTICULAR OBJECT RETRIEVAL WITH INTEGRAL MAX-POOLING OF CNN ACTIVATIONS

相关开源地址：

<https://github.com/Relja/netvlad>

<https://github.com/uzh-rpg/netvlad_tf_open>

<https://github.com/nmslib/hnswlib>

<https://github.com/facebookresearch/faiss>

<https://github.com/FALCONN-LIB/FALCONN>

1. 物体跟踪（Object Tracking）
2. 光流法
3. 卡尔曼滤波器
4. 均值漂移

物体跟踪在OpenCV里面都有实现，大多都是针对刚性物体，对于人脸这种物体不适合。

1. 物体分割（Object Segmentation）

目前主流的都是基于神经网络的。

FCN、SegNet、PSPNet、MaskRCNN 、DeepLab系列、RefineNet、DeeperLab

相关论文：

1. Fully Convolutional Networks for Semantic Segmentation
2. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation
3. Pyramid Scene Parsing Network
4. Mask R-CNN
5. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs
6. Rethinking Atrous Convolution for Semantic Image Segmentation
7. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation
8. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation
9. DeeperLab: Single-Shot Image Parser
10. MobileNetV2: Inverted Residuals and Linear Bottlenecks

相关开源地址：

<https://github.com/shekkizh/FCN.tensorflow>

<https://github.com/alexgkendall/caffe-segnet>

<https://github.com/hszhao/PSPNet>

<https://github.com/Vladkryvoruchko/PSPNet-Keras-tensorflow>

<https://github.com/matterport/Mask_RCNN>

<https://github.com/sthalles/deeplab_v3>

<https://github.com/DrSleep/tensorflow-deeplab-resnet>

<https://github.com/guosheng/refinenet>

<https://github.com/DrSleep/light-weight-refinenet>

1. 人脸检测（Face Detection）

传统方法：特征提取+分类器的方式

特征主要有HOG、HAAR等，分类器有Adaboost、SVM、Cascade等。

常用的开源库有：OpenCV、Dlib等。

深度学习：

MTCNN、PyramidBox、HR、Face R-CNN、SSH、RSA、S3FD、FaceBoxes

相关论文：

1. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks
2. PyramidBox: A Context-assisted Single Shot Face Detector.
3. Finding Tiny Faces
4. Face R-CNN
5. SSH: Single Stage Headless Face Detector
6. Recurrent Scale Approximation for Object Detection in CNN
7. S 3FD: Single Shot Scale-invariant Face Detector
8. FaceBoxes: A CPU Real-time Face Detector with High Accuracy

相关开源地址：

<https://github.com/kpzhang93/MTCNN_face_detection_alignment>

<https://github.com/EricZgw/PyramidBox>

<https://github.com/cydonia999/Tiny_Faces_in_Tensorflow>

<https://github.com/mahyarnajibi/SSH>

<https://github.com/sciencefans/RSA-for-object-detection>

<https://github.com/louis-she/sfd.pytorch>

<https://github.com/sfzhang15/FaceBoxes>

1. 人脸关键点对齐（Face Alignment）

一些人脸检测算法中会集成有人脸关键点对齐，在训练时2个任务的误差函数加权相加。对齐有2D和3D的区别，2D只考虑二维信息，3D需要有3维模型，能预测人脸的姿态信息。

2D关键点对齐：DCNN、MTCNN、TCDCN、LAB

3D关键点对齐：3DDFA、DenseReg、FAN、PRNet、PIPA

相关论文：

1. Facial Landmark Detection by Deep Multi-task Learning
2. Deep Convolutional Network Cascade for Facial Point Detection
3. Look at Boundary: A Boundary-Aware Face Alignment Algorithm
4. Face Alignment Across Large Poses: A 3D Solution
5. Pose-Invariant Face Alignment via CNN-Based Dense 3D Model Fitting
6. Dense Face Alignment
7. DenseReg: Fully Convolutional Dense Shape Regression In-the-Wild
8. How far are we from solving the 2D & 3D Face Alignment problem
9. Learning Dense Facial Correspondences in Unconstrained Images
10. Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network
11. Dense Face Alignment

相关开源地址：

<https://github.com/zhzhanp/TCDCN-face-alignment>

<https://github.com/wywu/LAB>

<https://github.com/cleardusk/3DDFA>

<https://github.com/ralpguler/DenseReg>

<https://github.com/YadiraF/PRNet>

<http://cvlab.cse.msu.edu/project-pifa.html>

1. 人脸识别（Face Recognition）

非神经网络：[GaussianFace](https://github.com/jangerritharms/GaussianFace)高斯脸

深度学习：大多数和损失函数设计有关

DeepFace、DeepID系列、VGGFace、FaceNet、CenterLoss、MarginalLoss、SphereFace、ArcFace

1、Surpassing Human-Level Face Verification Performance on LFW with GaussianFace

2、DeepFace: Closing the Gap to Human-Level Performance in Face Verification

3、Deep Learning Face Representation from Predicting 10,000 Classes

4、Deep Learning Face Representation by Joint Identification-Verification

5、DeepID3: Face Recognition with Very Deep Neural Networks

6、Deep Face Recognition

7、FaceNet: A Unified Embedding for Face Recognition and Clustering

8、A Discriminative Feature Learning Approach for Deep Face Recognition

9、Marginal Loss for Deep Face Recognition

10、SphereFace: Deep Hypersphere Embedding for Face Recognition

11、ArcFace: Additive Angular Margin Loss for Deep Face Recognition

相关开源地址:

<https://github.com/jangerritharms/GaussianFace>

<http://www.robots.ox.ac.uk/~vgg/software/vgg_face/>

<https://github.com/davidsandberg/facenet>

<https://github.com/wy1iu/sphereface>

<https://github.com/xialuxi/arcface-caffe>

1. 人像重建（Face Reconstruct）

基本上都是基于3D的，人像重建后可以进行姿态估计，以及换脸。有的换脸算法需要多张人脸训练GAN网络。

PRNet、VRN、Face2Face

1. State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications
2. 3D Face Reconstruction with Geometry Details from a Single Image
3. Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network
4. CNN-based Real-time Dense Face Reconstruction with Inverse-rendered Photo-realistic Face Images
5. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression
6. Deep Video Portraits
7. VDub: Modifying Face Video of Actors for Plausible Visual Alignment to a Dubbed Audio Track
8. paGAN: Real-time Avatars Using Dynamic Textures
9. On Face Segmentation, Face Swapping, and Face Perception

相关开源地址:

<https://github.com/YadiraF/PRNet>

<https://github.com/AaronJackson/vrn>

<https://github.com/deepfakes/faceswap>

<https://github.com/datitran/face2face-demo>

<https://github.com/YuvalNirkin/face_swap>

1. OCR字符识别

OCR涉及到字符场景定位和分割，以及字符识别。传统的方法是采用垂直方向直方图形式对字符进行分割，然后一个个字符分别送入分类器进行识别。由于CTC动态规划算法的出现，当今的主流模型是LSTM+CTC，采用和语音识别类似的自动语素分割的方式。检测框一般是水平的，如果要纠正还需要用Hough变换把文本方向纠正。

字符区域检测：

CTPN、TextBoxes++、AdvancedEast

相关论文：

1、Detecting Text in Natural Image with Connectionist Text Proposal Network

2、Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes

3、Single Shot Scene Text Retrieval

4、EAST: An Efficient and Accurate Scene Text Detector

### 5、DeepTextSpotter: An End-to-End Trainable Scene Text Localization and Recognition Framework

6、Recursive Recurrent Nets with Attention Modeling for OCR in the Wild

7、Multi-Oriented Text Detection with Fully Convolutional Networks

8、Accurate Text Localization in Natural Image with Cascaded Convolutional Text Network

9、总结Overview：

<https://github.com/whitelok/image-text-localization-recognition>

字符识别：

CRNN、GRCNN

1. Gated Recurrent Convolution Neural Network for OCR
2. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition

相关开源地址：

<https://github.com/eragonruan/text-detection-ctpn>

<https://github.com/MhLiao/TextBoxes_plusplus>

<https://github.com/lluisgomez/single-shot-str>

<https://github.com/huoyijie/AdvancedEAST>

<https://github.com/MichalBusta/DeepTextSpotter>

<https://github.com/Jianfeng1991/GRCNN-for-OCR>

1. 语音识别（Automatic Speech Recognition/Speech to Text）

传统方式基于GMM-HMM模型和Vertibi算法

深度学习：对WAV进行MFCC短时频谱信号提取，依次采用CNN卷积网络和LSTM循环网络以及CTC Loss误差函数进行建模。

GRU-CTC、DFCNN、DFSMN、DeepSpeech、CLDNN

相关论文

1. DEEP-FSMN FOR LARGE VOCABULARY CONTINUOUS SPEECH RECOGNITION
2. Deep Speech: Scaling up end-to-end speech recognition
3. CONVOLUTIONAL, LONG SHORT-TERM MEMORY, FULLY CONNECTED DEEP NEURAL NETWORKS

相关开源地址：

<https://github.com/buriburisuri/speech-to-text-wavenet>

<https://github.com/Kyubyong/tacotron>

<https://github.com/PaddlePaddle/DeepSpeech>

1. 说话人识别（Speaker Recognition/Identification/Verification）

目前深度学习并没有从根本上打败传统方法，但基于d-vector、x-vector的模型和TE2E/GE2E等的损失函数设计在短时长上比较占优势。

声纹识别的主要问题在于语音时长、文本无关、开集比对、背景噪声等问题上。

传统方法的state-of-the-art是i-vector，采用pLDA信道补偿算法，以前的方法有UBM-GMM和JFA信道补偿，但是需要大量的不同信道的语料样本。

传统方法的相关开源框架有Kaldi、ALIZE、SIDEKIT、pyannote-audio等。

深度学习的方法有d-vector、x-vector、j-vector（文本有关）以及结合E2E损失函数的模型。

相关开源地址：

<http://www-lium.univ-lemans.fr/sidekit/>

<https://alize.univ-avignon.fr/>

<http://www.kaldi-asr.org/>

<https://github.com/rajathkmp/speaker-verification>

<https://github.com/wangleiai/dVectorSpeakerRecognition>

<https://github.com/Janghyun1230/Speaker_Verification>

<https://github.com/pyannote/pyannote-audio>

1. 说话人语音分割（Speaker Diarization）

语音智能分割是基于说话人识别的，说话人识别效果的好坏决定语音分割的效果，当然还有切换点的识别效果也很重要。首先需要用VAD静音检测对语音进行分割，最简单的是用振幅来判断，如果有背景音则需要设计其他的VAD算法。切换点的判断可以通过BIC贝叶斯准则，最后就是聚类，判断哪些片段属于一个说话人，对于无监督学习算法，先验信息说话人数量显得尤为重要。

目前基于深度学习的框架也有不少，比如最近Google出的UIS-RNN。

相关论文：

1. FULLY SUPERVISED SPEAKER DIARIZATION
2. SPAKER DIARIZATION WITH LSTM

相关开源地址：

<https://github.com/google/uis-rnn>

<https://github.com/wq2012/SpectralCluster>

1. 语音合成（Text To Speech）

文本转语音，传统方法是采用语素拼接，这种方式合成的语音比较生硬，没有语调。当前Baidu、Google、FaceBook等出了很多基于深度学习的方法。一般的流程是先Encoder再Decoder，最后用Griffin-Lim算法或者WaveNet自回归模型将MFCC变成wave信号。

WaveNet系列（MFCC🡪WAVE）、DeepVoice系列、Tacotron系列、VoiceLoop、ClariNet

相关论文：

1. VOICELOOP: VOICE FITTING AND SYNTHESIS VIA A PHONOLOGICAL LOOP
2. TACOTRON: TOWARDS END-TO-END SPEECH SYNTHESIS
3. NATURAL TTS SYNTHESIS BY CONDITIONING WAVENET ON MEL SPECTROGRAM PREDICTIONS
4. Deep Voice: Real-time Neural Text-to-Speech
5. Deep Voice 2: Multi-Speaker Neural Text-to-Speech
6. DEEP VOICE 3: 2000-SPEAKER NEURAL TEXT-TO-SPEECH
7. WAVENET: A GENERATIVE MODEL FOR RAW AUDIO
8. Parallel WaveNet: Fast High-Fidelity Speech Synthesis
9. ClariNet: Parallel Wave Generation in End-to-End Text-to-Speech
10. SAMPLE EFFICIENT ADAPTIVE TEXT-TO-SPEECH

相关开源地址：

<https://github.com/ibab/tensorflow-wavenet>

<https://github.com/keithito/tacotron>

<https://github.com/Kyubyong/tacotron>

<https://github.com/c1niv/Voiceloop_TensorFlow>

<https://github.com/israelg99/deepvoice>

<https://github.com/andabi/parallel-wavenet-vocoder>

1. 声纹转换（Voice Conversion）

声纹转换其实就是TTS的多人版，根据说话人的不同将文本生成不同的wave信号。大多数都是在网络架构中加入说话人Embedding向量，如DeepVoice2/DeepVoice3，Tacotron2，有的甚至会在声码器Vocoder中加入，比如WaveNet。

相关开源地址：

<https://github.com/r9y9/deepvoice3_pytorch>

<https://github.com/Kyubyong/deepvoice3>

<https://github.com/Rayhane-mamah/Tacotron-2>

<https://github.com/GSByeon/multi-speaker-tacotron-tensorflow>