【语音识别】从入门到精通——最全干货大合集！

专知 专知 (/users/10000031042) 2018-11-07 11:52

关注文章

语音识别 (/articles?tagId=614)

【导读】语音识别领域最全入门资料、论文、代码、产品大合集！包括语音识别，语音合成，声纹识别等内容，一文在手，带你走进语音识别的世界。

作者 | 刘斌

编辑 | Xiaowen

入门学习

语音识别研究的四大前沿方

https://blog.csdn.net/haima1998/article/details/79094341 (https://blog.csdn.net/haima1998/article/details/79094341)

深度学习入门论文（语音识别领域）

https://blog.csdn.net/youyuyixiu/article/details/53764218 (https://blog.csdn.net/youyuyixiu/article/details/53764218)

论语音识别三大关键技术

https://blog.csdn.net/qq_34231800/article/details/80189617 (https://blog.csdn.net/qq_34231800/article/details/80189617)

深度学习与语音识别—常用声学模型简介

https://blog.csdn.net/dujiajiyi_xue5211314/article/details/53943313
(https://blog.csdn.net/dujiajiyi_xue5211314/article/details/53943313)

有趣的开源软件：语音识别工具Kaldi

https://blog.csdn.net/AMDS123/article/details/70313780 (https://blog.csdn.net/AMDS123/article/details/70313780)

神经网络-CNN结构和语音识别应用

https://blog.csdn.net/xmdxcsj/article/details/54695995 (https://blog.csdn.net/xmdxcsj/article/details/54695995)

语音识别概述

https://blog.csdn.net/shichaog/article/details/72528637 (https://blog.csdn.net/shichaog/article/details/72528637)

端到端语音识别

https://blog.csdn.net/xmdxcsj/article/details/70300546 (https://blog.csdn.net/xmdxcsj/article/details/70300546)

Attention在语音识别中的应用

https://blog.csdn.net/quheDiegooo/article/details/76842201 (https://blog.csdn.net/quheDiegooo/article/details/76842201)

语音合成技术

https://blog.csdn.net/wja8a45TJ1Xa/article/details/78599509?locationNum=8&fps=1
(https://blog.csdn.net/wja8a45TJ1Xa/article/details/78599509?locationNum=8&fps=1)

深度学习于语音合成研究综述

https://blog.csdn.net/weixin_37598106/article/details/81513816
(https://blog.csdn.net/weixin_37598106/article/details/81513816)

端到端的TTS深度学习模型tacotron(中文语音合成)

https://blog.csdn.net/yunnangf/article/details/79585089 (https://blog.csdn.net/yunnangf/article/details/79585089)

TACOTRON:端到端的语音合成

https://blog.csdn.net/Left_Think/article/details/74905928 (https://blog.csdn.net/Left_Think/article/details/74905928)

声纹识别技术简介

https://www.cnblogs.com/wuxian11/p/6498699.html (https://www.cnblogs.com/wuxian11/p/6498699.html)

声纹识别技术的现状、局限与趋势

https://blog.csdn.net/jojozhangju/article/details/78637221 (https://blog.csdn.net/jojozhangju/article/details/78637221)

声纹识别

https://www.jianshu.com/p/513dadeef1fd (https://www.jianshu.com/p/513dadeef1fd)

Deep speaker介绍

https://blog.csdn.net/Lauyeed/article/details/79936632 (https://blog.csdn.net/Lauyeed/article/details/79936632)

论文

语音识别 DNN

Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition(2012), George E. Dahl et al.

https://ieeexplore.ieee.org/document/5740583/?part=1 (https://ieeexplore.ieee.org/document/5740583/?part=1)

Deep Neural Networks for Acoustic Modeling in Speech Recognition(2012), Geoffrey Hinton et al.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6296526 (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6296526)

语音识别 CNN

Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition(2012), Ossama Abdel-Hamid et al.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6288864 (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6288864)

Deep convolutional neural networks for LVCSR(2013), Tara N. Sainath et al.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6639347 (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6639347)

Analysis of CNN-based speech recognition system using raw speech as input(2015), Dimitri Palaz et al.

https://infoscience.epfl.ch/record/210029/files/Palaz_INTERSPEECH_2015.pdf (https://infoscience.epfl.ch/record/210029/files/Palaz_INTERSPEECH_2015.pdf)

Very Deep Convolutional Neural Networks for Noise Robust Speech Recognition(2016), Yanmin Qian et al.

https://pdfs.semanticscholar.org/8043/cbfed66c98d2255ea79254de620837478099.pdf (https://pdfs.semanticscholar.org/8043/cbfed66c98d2255ea79254de620837478099.pdf)

Very deep multilingual convolutional neural networks for LVCSR(2016), Tom Sercu et al.

https://arxiv.org/pdf/1509.08967.pdf (https://arxiv.org/pdf/1509.08967.pdf)

Advances in Very Deep Convolutional Neural Networks for LVCSR(2016), Tom Sercu et al.

https://arxiv.org/pdf/1604.01792.pdf (https://arxiv.org/pdf/1604.01792.pdf)

Deep Convolutional Neural Networks with Layer-Wise Context Expansion and Attention(2016), Dong Yu et al.

https://pdfs.semanticscholar.org/716e/60cbbdacf01b3148e91a555358a96308b770.pdf?_ga=2.38333155.198966451.1540996486-1278087525.1535180761 (https://pdfs.semanticscholar.org/716e/60cbbdacf01b3148e91a555358a96308b770.pdf?_ga=2.38333155.198966451.1540996486-1278087525.1535180761)

语音识别 LSTM

Long short-term memory recurrent neural network architectures for large scale acoustic modeling(2014), Hasim Sak et al.

https://static.googleusercontent.com/media/research.google.com/zh-CN//pubs/archive/43905.pdf (https://static.googleusercontent.com/media/research.google.com/zh-CN//pubs/archive/43905.pdf)

Deep LSTM for Large Vocabulary Continuous Speech Recognition(2017), Xu Tian et al.

https://arxiv.org/pdf/1703.07090.pdf (https://arxiv.org/pdf/1703.07090.pdf)

English Conversational Telephone Speech Recognition by Humans and Machines(2017), George Saon et al.

https://arxiv.org/pdf/1703.02136.pdf (https://arxiv.org/pdf/1703.02136.pdf)

语音识别 CTC

Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks(2006), Alex Graves et al.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.6306&rep=rep1&type=pdf (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.6306&rep=rep1&type=pdf)

Towards End-to-End Speech Recognition with Recurrent Neural Networks(2014), Alex Graves et al.

http://proceedings.mlr.press/v32/graves14.pdf (http://proceedings.mlr.press/v32/graves14.pdf)

First-Pass Large Vocabulary Continuous Speech Recognition using Bi-Directional Recurrent DNNs(2014), Andrew L. Maas et al.

https://arxiv.org/pdf/1408.2873.pdf (https://arxiv.org/pdf/1408.2873.pdf)

Deep Speech: Scaling up end-to-end speech recognition(2014), Awni Y. Hannun et al.

https://arxiv.org/pdf/1412.5567.pdf (https://arxiv.org/pdf/1412.5567.pdf)

Online Sequence Training of Recurrent Neural Networks with Connectionist Temporal Classification(2015), Kyuyeon Hwang et al.

https://arxiv.org/pdf/1511.06841.pdf (https://arxiv.org/pdf/1511.06841.pdf)

Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition(2015), Hasim Sak et al.

https://arxiv.org/pdf/1507.06947.pdf (https://arxiv.org/pdf/1507.06947.pdf)

Joint CTC-Attention based End-to-End Speech Recognition using Multi-task Learning(2016), Suyoun Kim et al.

https://arxiv.org/pdf/1609.06773.pdf (https://arxiv.org/pdf/1609.06773.pdf)

Deep Speech 2: End-to-End Speech Recognition in English and Mandarin(2016), Dario Amodei et al.

http://proceedings.mlr.press/v48/amodei16.pdf (http://proceedings.mlr.press/v48/amodei16.pdf)

Wav2Letter: an End-to-End ConvNet-based Speech Recognition System(2016), Ronan Collobert et al.

https://arxiv.org/pdf/1609.03193.pdf (https://arxiv.org/pdf/1609.03193.pdf)

Multi-task Learning with CTC and Segmental CRF for Speech Recognition(2017), Liang Lu et al.

https://arxiv.org/pdf/1702.06378.pdf (https://arxiv.org/pdf/1702.06378.pdf)

Residual Convolutional CTC Networks for Automatic Speech Recognition(2017), Yisen Wang et al.`

https://arxiv.org/pdf/1702.07793.pdf (https://arxiv.org/pdf/1702.07793.pdf)

语音识别 Sequence Transduction

Sequence Transduction with Recurrent Neural Networks(2012), Alex Graves et al.

https://arxiv.org/pdf/1211.3711.pdf (https://arxiv.org/pdf/1211.3711.pdf)

语音识别 attention

End-to-end Continuous Speech Recognition using Attention-based Recurrent NN: First Results(2014), Jan Chorowski et al.

https://arxiv.org/pdf/1412.1602.pdf (https://arxiv.org/pdf/1412.1602.pdf)

Attention-Based Models for Speech Recognition(2015), Jan Chorowski et al.

https://arxiv.org/pdf/1506.07503.pdf (https://arxiv.org/pdf/1506.07503.pdf)

End-to-end attention-based large vocabulary speech recognition(2016), Dzmitry Bahdanau et al.

https://arxiv.org/pdf/1508.04395.pdf (https://arxiv.org/pdf/1508.04395.pdf)

Listen, attend and spell: A neural network for large vocabulary conversational speech recognition(2016), William Chan et al.

https://arxiv.org/pdf/1508.01211.pdf (https://arxiv.org/pdf/1508.01211.pdf)

End-to-end attention-based distant speech recognition with Highway LSTM(2016), Hassan Taherian.

https://arxiv.org/pdf/1610.05361.pdf (https://arxiv.org/pdf/1610.05361.pdf)

Direct Acoustics-to-Word Models for English Conversational Speech Recognition(2017), Kartik Audhkhasi et al.

https://arxiv.org/pdf/1703.07754.pdf (https://arxiv.org/pdf/1703.07754.pdf)

语音识别 多通道

Multichannel Signal Processing With Deep Neural Networks for Automatic Speech Recognition(2017), Tara N. Sainath et al.

http://www.ee.columbia.edu/~ronw/pubs/taslp2017-multichannel.pdf (http://www.ee.columbia.edu/~ronw/pubs/taslp2017-multichannel.pdf)

Multichannel End-to-end Speech Recognition(2017), Tsubasa Ochiai et al.

https://arxiv.org/pdf/1703.04783.pdf (https://arxiv.org/pdf/1703.04783.pdf)

语音合成 SampleRNN

SampleRNN: An Unconditional End-to-End Neural Audio Generation Model(2016), Soroush Mehri et al.

https://arxiv.org/pdf/1612.07837.pdf (https://arxiv.org/pdf/1612.07837.pdf)

语音合成 WaveNet

WaveNet: A Generative Model for Raw Audio(2016), Aäron van den Oord et al.

https://arxiv.org/pdf/1609.03499.pdf (https://arxiv.org/pdf/1609.03499.pdf)

语音合成 Deep Voice

Deep Voice: Real-time Neural Text-to-Speech(2017), Sercan O. Arik et al.

https://arxiv.org/pdf/1702.07825.pdf (https://arxiv.org/pdf/1702.07825.pdf)

语音合成 Deep Voice 2

Deep Voice 2: Multi-Speaker Neural Text-to-Speech(2017), Sercan Arik et al.

https://arxiv.org/pdf/1705.08947.pdf (https://arxiv.org/pdf/1705.08947.pdf)

语音合成 Tacotron

Tacotron: Towards End-to-End Speech Synthesis(2017), Yuxuan Wang et al.

https://pdfs.semanticscholar.org/f258/f0d3260e7fbdd961993086aaafa2afc714c9.pdf
(https://pdfs.semanticscholar.org/f258/f0d3260e7fbdd961993086aaafa2afc714c9.pdf)

语音合成 Tacotron 2

Natural tts synthesis by conditioning wavenet on mel spectrogram predictions(2018), Jonathan Shen et al.

https://sigport.org/sites/default/files/docs/ICASSP%202018%20-%20Tacotron%202.pdf
(https://sigport.org/sites/default/files/docs/ICASSP%202018%20-%20Tacotron%202.pdf)

语音合成 Voiceloop

Voiceloop: Voice Fitting and Synthesis via a Phonological Loop(2018), Yaniv Taigman et al.

https://arxiv.org/pdf/1707.06588.pdf (https://arxiv.org/pdf/1707.06588.pdf)

声纹识别 x-vector 使用TDNN提取语音的embedding

Deep Neural Network Embeddings for Text-Independent Speaker Verification(2017), David Snyder et al.

http://danielpovey.com/files/2017_interspeech_embeddings.pdf (http://danielpovey.com/files/2017_interspeech_embeddings.pdf)

百度 端到端声纹识别 Triplet Loss

Deep Speaker: an End-to-End Neural Speaker Embedding System(2017), Chao Li et al.

https://arxiv.org/pdf/1705.02304.pdf (https://arxiv.org/pdf/1705.02304.pdf)

声纹识别 3D卷积网络

Text-independent speaker verification using 3d convolutional neural networks(2018), Amirsina Torfi et al.

https://arxiv.org/pdf/1705.09422.pdf (https://arxiv.org/pdf/1705.09422.pdf)

声纹识别 端到端 GE2E

Generalized End-to-End Loss for Speaker Verfication(2018) Wan L et al.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8462665 (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8462665)

代码

kaldi

使用广泛的语音工具包

https://github.com/kaldi-asr/kaldi (https://github.com/kaldi-asr/kaldi)

A TensorFlow implementation of Baidu's DeepSpeech architecture

语音识别 Baidu DeepSpeech TensorFlow实现

https://github.com/mozilla/DeepSpeech (https://github.com/mozilla/DeepSpeech)

Speech-to-Text-WaveNet : End-to-end sentence level English speech recognition based on DeepMind's WaveNet and tensorflow

语音识别 DeepMind's WaveNet TensorFlow实现

https://github.com/buriburisuri/speech-to-text-wavenet (https://github.com/buriburisuri/speech-to-text-wavenet)

End-to-end automatic speech recognition system implemented in TensorFlow.

端到端语音识别 TensorFlow实现

https://github.com/zzw922cn/Automatic_Speech_Recognition (https://github.com/zzw922cn/Automatic_Speech_Recognition)

A PyTorch Implementation of End-to-End Models for Speech-to-Text

端到端语音识别 PyTorch实现

https://github.com/awni/speech (https://github.com/awni/speech)

A PaddlePaddle implementation of DeepSpeech2 architecture for ASR.

语音识别 DeepSpeech2 PaddlePaddle实现

https://github.com/PaddlePaddle/DeepSpeech (https://github.com/PaddlePaddle/DeepSpeech)

A TensorFlow Implementation of Tacotron: A Fully End-to-End Text-To-Speech Synthesis Model

语音合成 Tacotron TensorFlow实现

https://github.com/Kyubyong/tacotron (https://github.com/Kyubyong/tacotron)

Tacotron 2 - PyTorch implementation with faster-than-realtime inference

语音合成 Tacotron2 PyTorch实现

https://github.com/NVIDIA/tacotron2 (https://github.com/NVIDIA/tacotron2)

Deep neural networks for voice conversion (voice style transfer) in Tensorflow

语音合成 Deep-voice TensorFlow实现

https://github.com/andabi/deep-voice-conversion (https://github.com/andabi/deep-voice-conversion)

A method to generate speech across multiple speakers

语音合成 facebook PyTorch实现

https://github.com/facebookresearch/loop (https://github.com/facebookresearch/loop)

Speaker embedding(verification and recognition) using Pytorch

声纹识别 PyTorch实现

https://github.com/qqueing/DeepSpeaker-pytorch (https://github.com/qqueing/DeepSpeaker-pytorch)

Deep Learning & 3D Convolutional Neural Networks for Speaker Verification

声纹识别 3D卷积 TensorFlow实现

https://github.com/astofi/3D-convolutional-speaker-recognition (https://github.com/astofi/3D-convolutional-speaker-recognition)

产品应用

百度语音官网

http://yuyin.baidu.com/ (http://yuyin.baidu.com/)

腾讯AI开放平台

https://ai.qq.com/product/aaiasr.shtml (https://ai.qq.com/product/aaiasr.shtml)

讯飞开放平台

https://xfyun.cn/services/voicedictation (https://xfyun.cn/services/voicedictation)

必应语音

https://azure.microsoft.com/zh-cn/services/cognitive-services/speech/ (https://azure.microsoft.com/zh-cn/services/cognitive-services/speech/)

作者简介

刘斌，中科院自动化所博士生，研究方向为鲁棒性声学建模。

文章被以下专辑收录

AI 关注
(/collections/475)

3844 关注
(/collections/466)

TensorFlow 关注
(/collections/463)

0 | 收藏 | 举报

评论 按时间倒序

支持Markdown和数学公式,公式格式：\\(...\\)或\\[...\\]

B I H 预览

你有什么看法?

还没有内容

联系我们

欢迎来到TinyMind。
关于TinyMind的内容或商务合作、网站建议，举报不良信息等均可联系我们。
TinyMind客服邮箱：support@tinymind.com.cn
TinyMind客服微信：tinymind01 工作时间：周一至周五10:00-18:30