

Hung-Yi Lee

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/7015262/publications.pdf>

Version: 2024-02-01

90
papers

1,962
citations

933447

10
h-index

839539

18
g-index

90
all docs

90
docs citations

90
times ranked

832
citing authors

#	ARTICLE	IF	CITATIONS
1	Temporal pattern attention for multivariate time series forecasting. Machine Learning, 2019, 108, 1421-1441.	5.4	425
2	Mockingjay: Unsupervised Speech Representation Learning with Deep Bidirectional Transformer Encoders. , 2020, , .		160
3	TERA: Self-Supervised Learning of Transformer Encoder Representation for Speech. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2351-2366.	5.8	126
4	Audio Word2Vec: Unsupervised Learning of Audio Segment Representations Using Sequence-to-Sequence Autoencoder. , 0, , .		104
5	Audio Albert: A Lite Bert for Self-Supervised Learning of Audio Representation. , 2021, , .		82
6	Spoken Content Retrieval Beyond Cascading Speech Recognition with Text Retrieval. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 1389-1420.	5.8	73
7	Meta Learning for End-To-End Low-Resource Speech Recognition. , 2020, , .		48
8	Distilhubert: Speech Representation Learning by Layer-Wise Distillation of Hidden-Unit Bert. , 2022, , .		36
9	Again-VC: A One-Shot Voice Conversion Using Activation Guidance and Adaptive Instance Normalization. , 2021, , .		34
10	Improving Conditional Sequence Generative Adversarial Networks by Stepwise Evaluation. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 788-798.	5.8	31
11	Towards Audio to Scene Image Synthesis Using Generative Adversarial Network. , 2019, , .		30
12	One-Shot Voice Conversion by Vector Quantization. , 2020, , .		29
13	Audio Word2vec: Sequence-to-Sequence Autoencoding for Unsupervised Learning of Audio Segmentation and Representation. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 1481-1493.	5.8	28
14	Fragmentvc: Any-To-Any Voice Conversion by End-To-End Extracting and Fusing Fine-Grained Voice Fragments with Attention. , 2021, , .		27
15	Segmental Audio Word2Vec: Representing Utterances as Sequences of Vectors with Applications in Spoken Term Detection. , 2018, , .		26
16	Adversarial Training of End-to-end Speech Recognition Using a Criticizing Language Model. , 2019, , .		23
17	Interactive Spoken Document Retrieval With Suggested Key Terms Ranked by a Markov Decision Process. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 632-645.	3.2	21
18	Semi-Supervised Spoken Language Understanding via Self-Supervised Speech and Language Model Pretraining. , 2021, , .		21

#	ARTICLE	IF	CITATIONS
19	Investigating on Incorporating Pretrained and Learnable Speaker Representations for Multi-Speaker Multi-Style Text-to-Speech. , 2021, , .		21
20	Utilizing Self-Supervised Representations for MOS Prediction. , 0, , .		20
21	Improved spoken term detection by feature space pseudo-relevance feedback. , 0, , .		20
22	Towards Unsupervised Speech Recognition and Synthesis with Quantized Speech Representation Learning. , 2020, , .		18
23	Adversarial Defense for Automatic Speaker Verification by Cascaded Self-Supervised Learning Models. , 2021, , .		18
24	Improved spoken term detection with graph-based re-ranking in feature space. , 2011, , .		17
25	Enhanced Spoken Term Detection Using Support Vector Machines and Weighted Pseudo Examples. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 1272-1284.	3.2	17
26	Phonetic-and-Semantic Embedding of Spoken words with Applications in Spoken Content Retrieval. , 2018, , .		17
27	What Does a Network Layer Hear? Analyzing Hidden Representations of End-to-End ASR Through Speech Synthesis. , 2020, , .		17
28	Meta-TTS: Meta-Learning for Few-Shot Speaker Adaptive Text-to-Speech. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1558-1571.	5.8	17
29	ODSQA: Open-Domain Spoken Question Answering Dataset. , 2018, , .		16
30	Towards End-to-end Speech-to-text Translation with Two-pass Decoding. , 2019, , .		16
31	Self-Supervised Deep Learning for Fisheye Image Rectification. , 2020, , .		16
32	Unsupervised End-to-End Learning of Discrete Linguistic Units for Voice Conversion. , 0, , .		15
33	Completely Unsupervised Phoneme Recognition by Adversarially Learning Mapping Relationships from Audio Embeddings. , 0, , .		15
34	Defense for Black-Box Attacks on Anti-Spoofing Models by Self-Supervised Learning. , 0, , .		15
35	Spoken Knowledge Organization by Semantic Structuring and a Prototype Course Lecture System for Personalized Learning. IEEE/ACM Transactions on Audio Speech and Language Processing, 2014, 22, 883-898.	5.8	14
36	Order-Free Learning Alleviating Exposure Bias in Multi-Label Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34, 6038-6045.	4.9	14

#	ARTICLE	IF	CITATIONS
37	Rhythm-Flexible Voice Conversion Without Parallel Data Using Cycle-GAN Over Phoneme Posteriorgram Sequences. , 2018, , .		13
38	Scalable Sentiment for Sequence-to-Sequence Chatbot Response with Performance Analysis. , 2018, , .		13
39	Gate Activation Signal Analysis for Gated Recurrent Neural Networks and its Correlation with Phoneme Boundaries. , 0, , .		13
40	Improved spoken term detection using support vector machines with acoustic and context features from pseudo-relevance feedback. , 2011, , .		11
41	Defending Your Voice: Adversarial Attack on Voice Conversion. , 2021, , .		11
42	Improved lattice-based spoken document retrieval by directly learning from the evaluation measures. , 2009, , .		10
43	Enhancing query expansion for semantic retrieval of spoken content with automatically discovered acoustic patterns. , 2013, , .		10
44	Improved Semantic Retrieval of Spoken Content by Document/Query Expansion with Random Walk Over Acoustic Similarity Graphs. IEEE/ACM Transactions on Audio Speech and Language Processing, 2014, 22, 80-94.	5.8	10
45	Mitigating the impact of speech recognition errors on chatbot using sequence-to-sequence model. , 2017, , .		10
46	Improving the Adversarial Robustness for Speaker Verification by Self-Supervised Learning. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 202-217.	5.8	10
47	Improved spoken term detection using support vector machines based on lattice context consistency. , 2011, , .		9
48	Unsupervised two-stage keyword extraction from spoken documents by topic coherence and support vector machine. , 2012, , .		9
49	Personalized language modeling by crowd sourcing with social network data for voice access of cloud applications. , 2012, , .		9
50	Integrating Recognition and Retrieval With Relevance Feedback for Spoken Term Detection. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 2095-2110.	3.2	9
51	Interactive Spoken Content Retrieval by Deep Reinforcement Learning. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 2447-2459.	5.8	9
52	Mitigating the Impact of Speech Recognition Errors on Spoken Question Answering by Adversarial Domain Adaptation. , 2019, , .		9
53	Improved semantic retrieval of spoken content by language models enhanced with acoustic similarity graph. , 2012, , .		8
54	Hierarchical attention model for improved machine comprehension of spoken content. , 2016, , .		8

#	ARTICLE	IF	CITATIONS
55	Improving Automatic Speech Recognition and Speech Translation via Word Embedding Prediction. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 93-105.	5.8	8
56	Partially Fake Audio Detection by Self-Attention-Based Fake Span Discovery. , 2022, , .		8
57	Improved open-vocabulary spoken content retrieval with word and subword lattices using acoustic feature similarity. Computer Speech and Language, 2014, 28, 1045-1065.	4.3	7
58	An iterative deep learning framework for unsupervised discovery of speech features and linguistic units with applications on spoken term detection. , 2015, , .		7
59	Sequence-to-Sequence Automatic Speech Recognition with Word Embedding Regularization and Fused Decoding. , 2020, , .		7
60	How Far Are We from Robust Voice Conversion: A Survey. , 2021, , .		7
61	Integrating recognition and retrieval with user feedback: A new framework for spoken term detection. , 2010, , .		6
62	Towards unsupervised semantic retrieval of spoken content with query expansion based on automatically discovered acoustic patterns. , 2013, , .		6
63	Personalizing universal recurrent neural network language model with user characteristic features by social network crowdsourcing. , 2015, , .		6
64	Understanding Self-Attention of Self-Supervised Audio Transformers. , 0, , .		6
65	Improved spoken term detection by discriminative training of acoustic models based on user relevance feedback. , 0, , .		6
66	Learning Phone Recognition From Unpaired Audio and Phone Sequences Based on Generative Adversarial Network. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 230-243.	5.8	6
67	Adversarial Sample Detection for Speaker Verification by Neural Vocoders. , 2022, , .		6
68	Spoken term detection from bilingual spontaneous speech using code-switched lattice-based structures for words and subword units. , 2009, , .		5
69	Interactive spoken content retrieval by extended query model and continuous state space Markov Decision Process. , 2013, , .		5
70	Unsupervised domain adaptation for spoken document summarization with structured support vector machine. , 2013, , .		5
71	Seeing and hearing too: Audio representation for video captioning. , 2017, , .		5
72	Machine Comprehension of Spoken Content: TOEFL Listening Test and Spoken SQuAD. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 1469-1480.	5.8	5

#	ARTICLE	IF	CITATIONS
73	Supervised spoken document summarization jointly considering utterance importance and redundancy by structured support vector machine. , 0, , .		5
74	Non-Autoregressive Mandarin-English Code-Switching Speech Recognition. , 2021, , .		5
75	Semantic query expansion and context-based discriminative term modeling for spoken document retrieval. , 2012, , .		4
76	Abstractive headline generation for spoken content by attentive recurrent neural networks with ASR error modeling. , 2016, , .		4
77	End-to-End Whispered Speech Recognition with Frequency-Weighted Approaches and Pseudo Whisper Pre-training. , 2021, , .		4
78	A framework integrating different relevance feedback scenarios and approaches for spoken term detection. , 2010, , .		3
79	Utterance-level latent topic transition modeling for spoken documents and its application in automatic summarization. , 2012, , .		3
80	Characterizing the Adversarial Vulnerability of Speech self-Supervised Learning. , 2022, , .		3
81	An initial attempt to improve spoken term detection by learning optimal weights for different indexing features. , 2010, , .		2
82	Lifting motion planning for humanoid robots. , 2014, , .		2
83	Language Transfer of Audio Word2Vec: Learning Audio Segment Representations Without Target Language Data. , 2018, , .		2
84	Don't Speak Too Fast: The Impact of Data Bias on Self-Supervised Speech Models. , 2022, , .		2
85	Personalized acoustic modeling by weakly supervised multi-task deep learning using acoustic tokens discovered from unlabeled data. , 2017, , .		1
86	Recurrent Neural Network based language modeling with controllable external Memory. , 2017, , .		1
87	Personalized word representations carrying personalized semantics learned from social network posts. , 2017, , .		1
88	Domain Independent Key Term Extraction from Spoken Content Based on Context and Term Location Information in the Utterances. , 2018, , .		1
89	Guest Editorial Special Issue on Adversarial Learning in Computational Intelligence. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 4, 414-416.	4.9	0
90	A Fully Integrated 1.7mW Attention-Based Automatic Speech Recognition Processor. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, 69, 4178-4182.	3.0	0