Deep Audio-Visual Speech Recognition

IEEE Transactions on Pattern Analysis and Machine Intelligenc
44, 8717-8727
DOI: 10.1109/tpami.2018.2889052

Citation Report

#	Article	IF	Citations
1	Development of Output Correction Methodology for Long Short Term Memory-Based Speech Recognition. Sustainability, 2019, 11, 4250.	1.6	14
2	Towards Automatic Face-to-Face Translation. , 2019, , .		24
3	Deep Audio-visual System for Closed-set Word-level Speech Recognition. , 2019, , .		1
4	Critical assessment of methods of protein structure prediction (CASP)â€”Round XIII. Proteins: Structure, Function and Bioinformatics, 2019, 87, 1011-1020.	1.5	380
5	State of Charge Estimation for Lithium-Ion Batteries Using Model-Based and Data-Driven Methods: A Review. IEEE Access, 2019, 7, 136116-136136.	2.6	366
6	Genetic programming for multiple-feature construction on high-dimensional classification. Pattern Recognition, 2019, 93, 404-417.	5.1	59
7	Learning Disentangled Representation in Latent Stochastic Models: A Case Study with Image Captioning. , 2019, , .		0
8	Classification of Sonar Targets in Air: A Neural Network Approach. Sensors, 2019, 19, 1176.	2.1	15
9	â€œIs This an Example Image?â€•â€“ Predicting the Relative Abstractness Level of Image and Text. Lecture Notes in Computer Science, 2019, , 711-725.	1.0	0
10	Listening While Speaking and Visualizing: Improving ASR Through Multimodal Chain. , 2019, , .		3
11	Locally Confined Modality Fusion Network With a Global Perspective for Multimodal Human Affective Computing. IEEE Transactions on Multimedia, 2020, 22, 122-137.	5.2	47
12	Cross-Domain Deep Visual Feature Generation for Mandarin Audioâ€“Visual Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 185-197.	4.0	16
13	Machine Learning Methods in Drug Discovery. Molecules, 2020, 25, 5277.	1.7	182
14	Survey on Deep Neural Networks in Speech and Vision Systems. Neurocomputing, 2020, 417, 302-321.	3.5	117
15	A Multimodal Facial Emotion Recognition Framework through the Fusion of Speech with Visible and Infrared Images. Multimodal Technologies and Interaction, 2020, 4, 46.	1.7	28
16	Construction of word level Tibetan Lip Reading Dataset. , 2020, , .		2
17	Automatic Classification of Text Complexity. Applied Sciences (Switzerland), 2020, 10, 7285.	1.3	14
18	Indoor Scene Change Captioning Based on Multimodality Data. Sensors, 2020, 20, 4761.	2.1	16

#	Article	IF	Citations
19	Audioâ€“Visual Speech Recognition Based on Dual Cross-Modality Attentions with the Transformer Model. Applied Sciences (Switzerland), 2020, 10, 7263.	1.3	10
20	ASR is All You Need: Cross-Modal Distillation for Lip Reading. , 2020, , .		50
21	Efficient machine learning algorithm for electroencephalogram modeling in brainâ€“computer interfaces. Neural Computing and Applications, 2022, 34, 9233-9243.	3.2	4
22	How to Teach DNNs to Pay Attention to the Visual Modality in Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 1052-1064.	4.0	19
23	Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. Information (Switzerland), 2020, 11, 128.	1.7	39
24	A Long Sequence Speech Perceptual Hashing Authentication Algorithm Based on Constant Q Transform and Tensor Decomposition. IEEE Access, 2020, 8, 34140-34152.	2.6	7
25	Data Augmentation for Deep Learning-Based Radio Modulation Classification. IEEE Access, 2020, 8, 1498-1506.	2.6	84
26	Lipreading with DenseNet and resBi-LSTM. Signal, Image and Video Processing, 2020, 14, 981-989.	1.7	29
27	DMMAN: A two-stage audioâ€“visual fusion framework for sound separation and event localization. Neural Networks, 2021, 133, 229-239.	3.3	13
28	AVMSN: An Audio-Visual Two Stream Crowd Counting Framework Under Low-Quality Conditions. IEEE Access, 2021, 9, 80500-80510.	2.6	14
29	Multichannel speech separation using hybrid GOMF and enthalpy-based deep neural networks. Multimedia Systems, 2021, 27, 271-286.	3.0	7
30	Multimodal Corpus Analysis of Autoblog 2020: Lecture Videos in Machine Learning. Lecture Notes in Computer Science, 2021, , 262-270.	1.0	2
31	Audio-Visual Deep Neural Network for Robust Person Verification. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1079-1092.	4.0	31
32	Audio-Visual Multi-Channel Integration and Recognition of Overlapped Speech. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2067-2082.	4.0	14
33	Audio to Video: Generating a Talking Fake Agent. Advances in Intelligent Systems and Computing, 2021, , 212-227.	0.5	1
34	End-To-End Lip Synchronisation Based on Pattern Classification. , 2021, , .		1
35	An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1368-1396.	4.0	111
36	Robust Audio-Visual Speech Recognition Based on Hybrid Fusion. , 2021, , .		6

#	Article	IF	Citations
37	Top of the Class: Mining Product Characteristics Associated with Crowdfunding Success and Failure of Home Robots. International Journal of Social Robotics, 2022, 14, 149-163.	3.1	5
38	A novel weight initialization with adaptive hyper-parameters for deep semantic segmentation. Multimedia Tools and Applications, 2021, 80, 21771-21787.	2.6	2
39	Improving the Recognition Performance of Lip Reading Using the Concatenated Three Sequence Keyframe Image Technique. Engineering, Technology & Applied Science Research, 2021, 11, 6986-6992.	0.8	6
40	Deep Audio-visual Learning: A Survey. International Journal of Automation and Computing, 2021, 18, 351-376.	4.5	64
41	Classification of Handwritten Chinese Numbers with Convolutional Neural Networks. , 2021, , .		1
42	Surface crack detection based on image stitching and transfer learning with pretrained convolutional neural network. Structural Control and Health Monitoring, 2021, 28, e2766.	1.9	19
43	SpeakingFaces: A Large-Scale Multimodal Dataset of Voice Commands with Visual and Thermal Video Streams. Sensors, 2021, 21, 3465.	2.1	33
44	Multimodal Classification of Parkinsonâ€™s Disease in Home Environments with Resiliency to Missing Modalities. Sensors, 2021, 21, 4133.	2.1	11
45	Stock Price Forecast Based on CNN-BiLSTM-ECA Model. Scientific Programming, 2021, 2021, 1-20.	0.5	17
46	Toward Language-independent Lip Reading: A Transfer Learning Approach. , 2021, , .		1
47	Machine Learning Force Fields: Recent Advances and Remaining Challenges. Journal of Physical Chemistry Letters, 2021, 12, 6551-6564.	2.1	58
48	Audio-Visual Information Fusion Using Cross-Modal Teacher-Student Learning for Voice Activity Detection in Realistic Environments. , 0, , .		2
49	End-to-End Audio-Visual Speech Recognition for Overlapping Speech. , 0, , .		0
51	Survey on Machine Learning in Speech Emotion Recognition and Vision Systems Using a Recurrent Neural Network (RNN). Archives of Computational Methods in Engineering, 2022, 29, 1753-1770.	6.0	46
52	Deep Learning and Reinforcement Learning for Autonomous Unmanned Aerial Systems: Roadmap for Theory toÂDeployment. Studies in Computational Intelligence, 2021, , 25-82.	0.7	6
53	Recent Progress in the CUHK Dysarthric Speech Recognition System. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2267-2281.	4.0	25
54	Adaptive Semantic-Spatio-Temporal Graph Convolutional Network for Lip Reading. IEEE Transactions on Multimedia, 2022, 24, 3545-3557.	5.2	9
55	Speaker-Dependent Visual Command Recognition in Vehicle Cabin: Methodology and Evaluation. Lecture Notes in Computer Science, 2021, , 291-302.	1.0	7

#	Article	IF	Citations
56	Using Deep Convolutional LSTM Networks for Learning Spatiotemporal Features. Lecture Notes in Computer Science, 2020, , 307-320.	1.0	4
57	Self-supervised Learning of Audio-Visual Objects from Video. Lecture Notes in Computer Science, 2020, , 208-224.	1.0	70
58	BSL-1K: Scaling Up Co-articulated Sign Language Recognition Using Mouthing Cues. Lecture Notes in Computer Science, 2020, , 35-53.	1.0	54
59	Multi-channel Transformers for Multi-articulatory Sign Language Translation. Lecture Notes in Computer Science, 2020, , 301-319.	1.0	43
60	Dynamic Facial Features in Positive-Emotional Speech for Identification of Depressive Tendencies. Smart Innovation, Systems and Technologies, 2020, , 127-134.	0.5	7
61	Multimodal Intelligence: Representation Learning, Information Fusion, and Applications. IEEE Journal on Selected Topics in Signal Processing, 2020, 14, 478-493.	7.3	182
62	A Lip Sync Expert Is All You Need for Speech to Lip Generation In the Wild. , 2020, , .		244
63	TieLent. , 2020, , .		11
64	Information cascades prediction with attention neural network. Human-centric Computing and Information Sciences, 2020, 10, .	6.1	14
65	Multimodal and Multiresolution Speech Recognition with Transformers. , 2020, , .		22
66	CroMM-VSR: Cross-Modal Memory Augmented Visual Speech Recognition. IEEE Transactions on Multimedia, 2022, 24, 4342-4355.	5.2	11
67	Real Time Online Visual End Point Detection Using Unidirectional LSTM. , 0, , .		2
68	Spotting Visual Keywords from Temporal Sliding Windows. , 2019, , .		3
69	Audiovisual Transformer Architectures for Large-Scale Classification and Synchronization of Weakly Labeled Audio Events. , 2019, , .		6
70	One Perceptron to Rule Them All: Language, Vision, Audio and Speech. , 2020, , .		0
71	A Survey of Lipreading Methods Based on Deep Learning. , 2020, , .		1
72	A methodology of multimodal corpus creation for audio-visual speech recognition in assistive transport systems. Informatization and Communication, 2020, 5, 87-93.	0.0	0
73	Phonemes Convey Embodied Emotion. , 2021, , 221-243.		2

#	Article	IF	Citations
74	SpotFast Networks with Memory Augmented Lateral Transformers for Lipreading. Communications in Computer and Information Science, 2020, , 554-561.	0.4	7
76	FastLR. , 2020, , .		7
77	Bimodal variational autoencoder for audiovisual speech recognition. Machine Learning, 2023, 112, 1201-1226.	3.4	4
78	A General Survey on Attention Mechanisms in Deep Learning. IEEE Transactions on Knowledge and Data Engineering, 2023, 35, 3279-3298.	4.0	92
79	Development of Visual and Audio Speech Recognition Systems Using Deep Neural Networks. , 2021, , .		1
80	Speech Reconstruction With Reminiscent Sound Via Visual Voice Memory. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 3654-3667.	4.0	10
81	Resource-Adaptive Deep Learning for Visual Speech Recognition. , 0, , .		3
82	Multi-Modal Embeddings Using Multi-Task Learning for Emotion Recognition. , 0, , .		9
83	BLSTM-Driven Stream Fusion for Automatic Speech Recognition: Novel Methods and a Multi-Size Window Fusion Example. , 0, , .		1
84	A Review on Generative Adversarial Networks. , 2020, , .		1
85	Audio Visual Speech Recognition using Feed Forward Neural Network Architecture. , 2020, , .		3
86	A Generative Answer Aggregation Model for Sentence-Level Crowdsourcing Tasks. IEEE Transactions on Knowledge and Data Engineering, 2023, 35, 3299-3312.	4.0	1
87	Selective Listening by Synchronizing Speech With Lips. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1650-1664.	4.0	9
88	A CNN-Based Method forÂAAPL Stock Price Trend Prediction Using Historical Data andÂTechnical Indicators. Smart Innovation, Systems and Technologies, 2022, , 25-33.	0.5	1
89	Combining audio and visual speech recognition using LSTM and deep convolutional neural network. International Journal of Information Technology (Singapore), 2022, 14, 3425-3436.	1.8	21
90	Visual Speech Recognition. International Journal of Advanced Research in Science, Communication and Technology, 0, , 355-358.	0.0	0
91	The Right to Talk: An Audio-Visual Transformer Approach. , 2021, , .		13
92	Ubi-SleepNet. , 2021, 5, 1-33.		2

#	Article	IF	Citations
93	Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation. Journal of King Saud University - Computer and Information Sciences, 2022, 34, 9710-9722.	2.7	4
94	A Comprehensive Review of Recent Automatic Speech Summarization and Keyword Identification Techniques. Learning and Analytics in Intelligent Systems, 2022, , 111-126.	0.5	15
95	Recent Advances in End-to-End Automatic Speech Recognition. APSIPA Transactions on Signal and Information Processing, 2022, 11, .	2.6	109
97	Audiovisual speech recognition for Kannada language using feed forward neural network. Neural Computing and Applications, 2022, 34, 15603-15615.	3.2	3
98	Masked Contrastive Representation Learning for Reinforcement Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, PP, 1-1.	9.7	9
99	End-to-End Lip-Reading Without Large-Scale Data. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 2076-2090.	4.0	4
100	Review on research progress of machine lip reading. Visual Computer, 2023, 39, 3041-3057.	2.5	2
101	Siamese decoupling network for speaker-independent lipreading. Journal of Electronic Imaging, 2022, 31, .	0.5	0
102	Reliability-Based Large-Vocabulary Audio-Visual Speech Recognition. Sensors, 2022, 22, 5501.	2.1	1
103	LipSound2: Self-Supervised Pre-Training for Lip-to-Speech Reconstruction and Lip Reading. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35, 2772-2782.	7.2	7
104	Joint modelling of audio-visual cues using attention mechanisms for emotion recognition. Multimedia Tools and Applications, 2023, 82, 11239-11264.	2.6	5
105	A Lip Reading Method Based on 3D Convolutional Vision Transformer. IEEE Access, 2022, 10, 77205-77212.	2.6	8
106	Importance-Aware Information Bottleneck Learning Paradigm for Lip Reading. IEEE Transactions on Multimedia, 2022, , 1-13.	5.2	1
107	Arbitrary Voice Conversion via Adversarial Learning and Cycle Consistency Loss. Lecture Notes in Computer Science, 2022, , 569-578.	1.0	0
108	Look&listen: Multi-Modal Correlation Learning for Active Speaker Detection and Speech Enhancement. IEEE Transactions on Multimedia, 2023, 25, 5800-5812.	5.2	7
109	Sub-word Level Lip Reading With Visual Attention. , 2022, , .		28
110	Cross-modal mask fusion and modality-balanced audio-visual speech recognition. , 2022, , .		1
111	Audio-video fusion strategies for active speaker detection in meetings. Multimedia Tools and Applications, 2023, 82, 13667-13688.	2.6	2

#	Article	IF	Citations
112	An Interference-Resistant and Low-Consumption Lip Recognition Method. Electronics (Switzerland), 2022, 11, 3066.	1.8	0
113	Synthesizing Talking Face Videos withÂaÂSpatial Attention Mechanism. Lecture Notes in Computer Science, 2022, , 519-528.	1.0	0
114	Speaker-Adaptive Lip Reading with User-Dependent Padding. Lecture Notes in Computer Science, 2022, , 576-593.	1.0	6
115	Learning Visual Styles fromÂAudio-Visual Associations. Lecture Notes in Computer Science, 2022, , 235-252.	1.0	4
116	Temporal andÂCross-modal Attention forÂAudio-Visual Zero-Shot Learning. Lecture Notes in Computer Science, 2022, , 488-505.	1.0	6
117	Visual speech recognition for multiple languages in the wild. Nature Machine Intelligence, 2022, 4, 930-939.	8.3	22
118	CroReLU: Cross-Crossing Space-Based Visual Activation Function for Lung Cancer Pathology Image Recognition. Cancers, 2022, 14, 5181.	1.7	5
119	Learning Contextually Fused Audio-Visual Representations For Audio-Visual Speech Recognition. , 2022, , .		2
120	Lipreading Using Liquid State Machine with STDP-Tuning. Applied Sciences (Switzerland), 2022, 12, 10484.	1.3	1
122	An online intelligent electronic medical record system via speech recognition. International Journal of Distributed Sensor Networks, 2022, 18, 155013292211344.	1.3	0
123	Combined Prediction Method of Short-Term Distance Headway Based on EB-GRA-TCN. Journal of Advanced Transportation, 2022, 2022, 1-12.	0.9	1
124	Audio-Visual Kinship Verification: A New Dataset and a Unified Adaptive Adversarial Multimodal Learning Approach. IEEE Transactions on Cybernetics, 2024, 54, 1523-1536.	6.2	1
125	Semantic and Relation Modulation for Audio-Visual Event Localization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45, 7711-7725.	9.7	5
126	Contrastive Positive Sample Propagation Along the Audio-Visual Event Line. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45, 7239-7257.	9.7	2
127	Vision-Guided Speaker Embedding Based Speech Separation. , 2022, , .		1
128	A comprehensive review of task understanding of command-triggered execution of tasks for service robots. Artificial Intelligence Review, 0, , .	9.7	0
129	Electromyogramâ€Based Lipâ€Reading via Unobtrusive Dry Electrodes and Machine Learning Methods. Small, 2023, 19, .	5.2	4
130	Lip Reading Bengali Words. , 2022, , .		0

#	Article	IF	Citations
131	Deep Learning Approach For Human Emotion-Gender-Age Recognition. , 2022, , .		0
132	Audio-Visual Overlapped Speech Detection for Spontaneous Distant Speech. IEEE Access, 2023, 11, 27426-27432.	2.6	0
133	EvIs-Kitchen: Egocentric Human Activities Recognition withÂVideo andÂInertial Sensor Data. Lecture Notes in Computer Science, 2023, , 373-384.	1.0	2
134	Improving Speech Recognition Performance in Noisy Environments by Enhancing Lip Reading Accuracy. Sensors, 2023, 23, 2053.	2.1	2
135	Audio-Visual Speech and Gesture Recognition by Sensors of Mobile Devices. Sensors, 2023, 23, 2284.	2.1	25
136	Voice Keyword Spotting on Edge Devices. , 2022, , .		1
137	Multimodal Sensor-Input Architecture with Deep Learning for Audio-Visual Speech Recognition in Wild. Sensors, 2023, 23, 1834.	2.1	2
138	Missing data in multi-omics integration: Recent advances through artificial intelligence. Frontiers in Artificial Intelligence, 0, 6, .	2.0	17
139	Improvement of Acoustic Models Fused with Lip Visual Information for Low-Resource Speech. Sensors, 2023, 23, 2071.	2.1	0
140	Temporal Sentence Grounding in Videos: A Survey and Future Directions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45, 10443-10465.	9.7	2
141	Audioâ€“visual keyword transformer for unconstrained sentenceâ€level keyword spotting. CAAI Transactions on Intelligence Technology, 2024, 9, 142-152.	3.4	2
142	End-to-End Chinese Lip-Reading Recognition Based on Multi-modal Fusion. , 2022, , .		1
143	Research on Robust Audio-Visual Speech Recognition Algorithms. Mathematics, 2023, 11, 1733.	1.1	2
144	Enhance Gesture Recognition viaÂVisual-Audio Modal Embedding. Lecture Notes in Computer Science, 2023, , 391-402.	1.0	0
145	Self-Supervised Training of Speaker Encoder With Multi-Modal Diverse Positive Pairs. IEEE/ACM Transactions on Audio Speech and Language Processing, 2023, 31, 1706-1719.	4.0	0
152	Dual-Path Cross-Modal Attention for Better Audio-Visual Speech Extraction. , 2023, , .		0
153	Imaginary Voice: Face-Styled Diffusion Model for Text-to-Speech. , 2023, , .		1
155	Lip Detection and Recognition-A Review1. , 2023, , .		0

#	Article	IF	Citations
159	Implementation of a Deepfake Detection System using Convolutional Neural Networks and Adversarial Training. , 2023, , .		0
164	Improving Audio-Visual Speech Recognition by Lip-Subword Correlation Based Visual Pre-training and Cross-Modal Fusion Encoder. , 2023, , .		0
167	MultiLingualSync: A Novel Method for Generating Lip-Synced Videos in Multiple Languages. , 2023, , .		0
168	Russian Language Speech Generation from Facial Video Recordings Using Variational Autoencoder. Studies in Computational Intelligence, 2023, , 489-498.	0.7	0
173	Emotional Speech-Driven Animation with Content-Emotion Disentanglement. , 2023, , .		1
174	Deep Learning for the Recognition of Skin Cancer. , 2023, , .		0
177	Utilizing Video Word Boundaries andÂFeature-Based Knowledge Distillation Improving Sentence-Level Lip Reading. Lecture Notes in Computer Science, 2024, , 269-281.	1.0	0
179	HDTR-Net: A Real-Time High-Definition Teeth Restoration Network forÂArbitrary Talking Face Generation Methods. Lecture Notes in Computer Science, 2024, , 89-103.	1.0	0
181	EMMN: Emotional Motion Memory Network for Audio-driven Emotional Talking Face Generation. , 2023, , .		0
182	MixSpeech: Cross-Modality Self-Learning with Audio-Visual Stream Mixup for Visual Speech Translation and Recognition. , 2023, , .		0
183	Lip2Vec: Efficient and Robust Visual Speech Recognition via Latent-to-Latent Visual to Audio Representation Mapping. , 2023, , .		0
184	Speech2Lip: High-fidelity Speech to Lip Generation by Learning from a Short Video. , 2023, , .		0
185	XVO: Generalized Visual Odometry via Cross-Modal Self-Training. , 2023, , .		0
187	Parameter-Efficient Cross-Language Transfer Learning for a Language-Modular Audiovisual Speech Recognition. , 2023, , .		0
188	Improving Audiovisual Active Speaker Detection in Egocentric Recordings with the Data-Efficient Image Transformer. , 2023, , .		0
189	Av-Data2Vec: Self-Supervised Learning of Audio-Visual Speech Representations with Contextualized Target Representations. , 2023, , .		0
191	Voice-Driven Virtual Population Type Method Based on Attention Mechanism. , 2023, , .		0
192	MAVAR-SE: Multi-scale Audio-Visual Association Representation Network for End-to-End Speaker Extraction. Lecture Notes in Computer Science, 2024, , 227-238.	1.0	0

#	Article	IF	Citations
193	Cycle-Consistent Generative Adversarial Network Architectures for Audio Visual Speech Recognition. , 2023, , .		0
195	Dip Into: A Novel Method for Visual Speech Recognition using Deep Learning. , 2023, , .		0
197	Multimodal active speaker detection using cross-attention and contextual information. , 2024, , .		0
199	Text-to-Feature Diffusion forÂAudio-Visual Few-Shot Learning. Lecture Notes in Computer Science, 2024, , 491-507.	1.0	0