

# Andrew Zisserman

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/6676312/publications.pdf>

Version: 2024-02-01

113  
papers

43,295  
citations

126708

33  
h-index

233125

45  
g-index

117  
all docs

117  
docs citations

117  
times ranked

22767  
citing authors

#	ARTICLE	IF	CITATIONS
1	LAEO-Net++: Revisiting People Looking at Each Other in Videos. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44, 3069-3081.	9.7	7
2	Scaling Up Sign Spotting Through Sign Language Dictionaries. International Journal of Computer Vision, 2022, 130, 1416-1439.	10.9	4
3	Synthetic Humans for Action Recognition from Unseen Viewpoints. International Journal of Computer Vision, 2021, 129, 2264-2287.	10.9	44
4	SeeHear: Signer Diarisation and a New Dataset. , 2021, , .		1
5	Slow-Fast Auditory Streams for Audio Recognition. , 2021, , .		24
6	QUERYD: A Video Dataset with High-Quality Text and Audio Narrations. , 2021, , .		6
7	Automated Video Labelling: Identifying Faces by Corroborative Evidence. , 2021, , .		7
8	Face, Body, Voice: Video Person-Clustering with Multiple Modalities. , 2021, , .		10
9	Automated audiovisual behavior recognition in wild primates. Science Advances, 2021, 7, eabi4883.	4.7	32
10	LSD-C: Linearly Separable Deep Clusters. , 2021, , .		6
11	TeachText: CrossModal Generalized Distillation for Text-Video Retrieval. , 2021, , .		65
12	Self-supervised Video Object Segmentation by Motion Grouping. , 2021, , .		56
13	Automated Video Face Labelling for Films and TV Material. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42, 780-792.	9.7	3
14	Semi-Supervised Learning with Scarce Annotations. , 2020, , .		17
15	End-to-End Learning of Visual Representations From Uncurated Instructional Videos. , 2020, , .		308
16	Sight to Sound: An End-to-End Approach for Visual Piano Transcription. , 2020, , .		14
17	Vggsound: A Large-Scale Audio-Visual Dataset. , 2020, , .		119
18	Identifying Scoliosis in Population-Based Cohorts: Automation of a Validated Method Based on Total Body Dual Energy X-ray Absorptiometry Scans. Calcified Tissue International, 2020, 106, 378-385.	1.5	11

#	ARTICLE	IF	CITATIONS
19	Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics. <i>Scientific Data</i> , 2020, 7, 102.	2.4	9
20	Self-supervised Learning of Audio-Visual Objects from Video. <i>Lecture Notes in Computer Science</i> , 2020, , 208-224.	1.0	70
21	Smooth-AP: Smoothing the Path Towards Large-Scale Image Retrieval. <i>Lecture Notes in Computer Science</i> , 2020, , 677-694.	1.0	62
22	Memory-Augmented Dense Predictive Coding for Video Representation Learning. <i>Lecture Notes in Computer Science</i> , 2020, , 312-329.	1.0	70
23	Amplifying Key Cues for Human-Object-Interaction Detection. <i>Lecture Notes in Computer Science</i> , 2020, , 248-265.	1.0	38
24	The Use and Reuse of Printed Illustrations in 15th-Century Venetian Editions. <i>Studi Di Storia</i> , 2020, , .	0.0	2
25	Utterance-level Aggregation for Speaker Recognition in the Wild. , 2019, , .		161
26	Chimpanzee face recognition from videos in the wild using deep learning. <i>Science Advances</i> , 2019, 5, eaaw0736.	4.7	127
27	You Said That?: Synthesising Talking Faces from Audio. <i>International Journal of Computer Vision</i> , 2019, 127, 1767-1779.	10.9	87
28	Count, Crop and Recognise: Fine-Grained Recognition in the Wild. , 2019, , .		4
29	Self-Supervised Learning of Class Embeddings from Video. , 2019, , .		13
30	EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition. , 2019, , .		156
31	LAEO-Net: Revisiting People Looking at Each Other in Videos. , 2019, , .		38
32	Video Representation Learning by Dense Predictive Coding. , 2019, , .		203
33	Non-contact physiological monitoring of preterm infants in the Neonatal Intensive Care Unit. <i>Npj Digital Medicine</i> , 2019, 2, 128.	5.7	82
34	Learning to Predict 3D Surfaces of Sculptures from Single and Multiple Views. <i>International Journal of Computer Vision</i> , 2019, 127, 1780-1800.	10.9	9
35	From Same Photo: Cheating on Visual Kinship Challenges. <i>Lecture Notes in Computer Science</i> , 2019, , 654-668.	1.0	8
36	Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. <i>Medical Image Analysis</i> , 2018, 46, 1-14.	7.0	72

#	ARTICLE	IF	CITATIONS
37	Crystal nucleation in metallic alloys using x-ray radiography and machine learning. Science Advances, 2018, 4, eaar4004.	4.7	74
38	Microscopy cell counting and detection with fully convolutional regression networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2018, 6, 283-292.	1.3	251
39	Seeing Voices and Hearing Faces: Cross-Modal Biometric Matching. , 2018, , .		117
40	Emotion Recognition in Speech using Cross-Modal Transfer in the Wild. , 2018, , .		180
41	VGGFace2: A Dataset for Recognising Faces across Pose and Age. , 2018, , .		1,475
42	Objects that Sound. Lecture Notes in Computer Science, 2018, , 451-466.	1.0	178
43	Learnable PINs: Cross-modal Embeddings for Person Identity. Lecture Notes in Computer Science, 2018, , 73-89.	1.0	65
44	Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Scientific Data, 2018, 5, 180124.	2.4	33
45	Learning to Read by Spelling. , 2018, , .		8
46	ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. European Spine Journal, 2017, 26, 1374-1383.	1.0	131
47	SpineNet: Automated classification and evidence visualization in spinal MRIs. Medical Image Analysis, 2017, 41, 63-73.	7.0	97
48	Template Adaptation for Face Verification and Identification. , 2017, , .		61
49	Multi-Task Convolutional Neural Network for Patient Detection and Skin Segmentation in Continuous Non-Contact Vital Sign Monitoring. , 2017, , .		48
50	Look, Listen and Learn. , 2017, , .		414
51	Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. , 2017, , .		4,575
52	Discovery of Rare Phenotypes in Cellular Images Using Weakly Supervised Deep Learning. , 2017, , .		4
53	Out of Time: Automated Lip Sync in the Wild. Lecture Notes in Computer Science, 2017, , 251-263.	1.0	120
54	From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV series without a Script. , 2017, , .		21

#	ARTICLE	IF	CITATIONS
55	Synthetic Data for Text Localisation in Natural Images. , 2016, , .		906
56	Convolutional Two-Stream Network Fusion for Video Action Recognition. , 2016, , .		1,749
57	Counting in the Wild. Lecture Notes in Computer Science, 2016, , 483-498.	1.0	127
58	3D Shape Attributes. , 2016, , .		15
59	Reading Text in the Wild with Convolutional Neural Networks. International Journal of Computer Vision, 2016, 116, 1-20.	10.9	831
60	Detecting overlapping instances in microscopy images using extremal region trees. Medical Image Analysis, 2016, 27, 3-16.	7.0	63
61	Flowing ConvNets for Human Pose Estimation in Videos. , 2015, , .		350
62	On-the-fly learning for visual search of large-scale image and video datasets. International Journal of Multimedia Information Retrieval, 2015, 4, 75-93.	3.6	31
63	Human pose search using deep poselets. , 2015, , .		6
64	Deep Face Recognition. , 2015, , .		2,584
65	Diagnostically relevant facial gestalt information from ordinary photos. ELife, 2014, 3, e02020.	2.8	129
66	Total Cluster. , 2014, , .		22
67	Triangulation Embedding and Democratic Aggregation for Image Search. , 2014, , .		167
68	Immediate, Scalable Object Category Detection. , 2014, , .		8
69	Automatic and Efficient Human Pose Estimation for Sign Language Videos. International Journal of Computer Vision, 2014, 110, 70-90.	10.9	50
70	Interactive Object Counting. Lecture Notes in Computer Science, 2014, , 504-518.	1.0	85
71	A Compact and Discriminative Face Track Descriptor. , 2014, , .		90
72	Seeing the Arrow of Time. , 2014, , .		50

#	ARTICLE	IF	CITATIONS
73	Deblurring Shaken and Partially Saturated Images. International Journal of Computer Vision, 2014, 110, 185-201.	10.9	140
74	Return of the Devil in the Details: Delving Deep into Convolutional Nets. , 2014, , .		1,522
75	Symbiotic Segmentation and Part Localization for Fine-Grained Categorization. , 2013, , .		170
76	All About VLAD. , 2013, , .		489
77	Learning to Detect Partially Overlapping Instances. , 2013, , .		29
78	Cats and dogs. , 2012, , .		518
79	Sparse kernel approximations for efficient classification and detection. , 2012, , .		63
80	On-the-fly specific person retrieval. , 2012, , .		22
81	Non-uniform Deblurring for Shaken Images. International Journal of Computer Vision, 2012, 98, 168-186.	10.9	390
82	Humanising GrabCut: Learning to segment humans using the Kinect. , 2011, , .		29
83	Deblurring shaken and partially saturated images. , 2011, , .		49
84	BiCoS: A Bi-level co-segmentation method for image classification. , 2011, , .		56
85	Smooth object retrieval using a bag of boundaries. , 2011, , .		52
86	Tabula rasa: Model transfer for object category detection. , 2011, , .		227
87	Geometric Latent Dirichlet Allocation on a Matching Graph for Large-scale Image Datasets. International Journal of Computer Vision, 2011, 95, 138-153.	10.9	51
88	Upper Body Detection and Tracking in Extended Signing Sequences. International Journal of Computer Vision, 2011, 95, 180-197.	10.9	52
89	Learning equivariant structured output SVM regressors. , 2011, , .		23
90	The devil is in the details: an evaluation of recent feature encoding methods. , 2011, , .		595

#	ARTICLE	IF	CITATIONS
91	Where Is the Archive?. Photography and Culture, 2010, 3, 331-342.	0.2	1
92	The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 2010, 88, 303-338.	10.9	12,097
93	Illuminance Flow Estimation by Regression. International Journal of Computer Vision, 2010, 90, 304-312.	10.9	0
94	Non-uniform deblurring for shaken images. , 2010, , .		174
95	Geodesic star convexity for interactive image segmentation. , 2010, , .		259
96	“Who are you?” - Learning person specific classifiers from video. , 2009, , .		137
97	“Who are you?” - Learning person specific classifiers from video. , 2009, , .		1
98	Learning an Alphabet of Shape and Appearance for Multi-Class Object Detection. International Journal of Computer Vision, 2008, 80, 16-44.	10.9	81
99	Lost in quantization: Improving particular object retrieval in large scale image databases. , 2008, , .		1,054
100	Object Mining Using a Matching Graph on Very Large Image Collections. , 2008, , .		64
101	Automated Flower Classification over a Large Number of Classes. , 2008, , .		1,529
102	An Exemplar Model for Learning Object Classes. , 2007, , .		193
103	Object retrieval with large vocabularies and fast spatial matching. , 2007, , .		2,144
104	Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval. , 2007, , .		630
105	Object Level Grouping for Video Shots. International Journal of Computer Vision, 2006, 67, 189-210.	10.9	71
106	A Statistical Approach to Texture Classification from Single Images. International Journal of Computer Vision, 2005, 62, 61-81.	10.9	34
107	A Statistical Approach to Texture Classification from Single Images. International Journal of Computer Vision, 2005, 62, 61-81.	10.9	998
108	Creating Architectural Models from Images. Computer Graphics Forum, 1999, 18, 39-50.	1.8	194

#	ARTICLE	IF	CITATIONS
109	Resolving ambiguities in auto-calibration. Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences, 1998, 356, 1193-1211.	1.6	44
110	A framework for spatiotemporal control in the tracking of visual contours. International Journal of Computer Vision, 1993, 11, 127-145.	10.9	300
111	VoxCeleb: A Large-Scale Speaker Identification Dataset. , 0, , .		786
112	VoxCeleb2: Deep Speaker Recognition. , 0, , .		964
113	Bounding an archiving: assessing the relative completeness of the Jacques Toussele archive using pattern-matching and face-recognition. Visual Studies, 0, , 1-25.	0.3	0