

# Song-Chun Zhu

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/3317673/publications.pdf>

Version: 2024-02-01

126  
papers

5,307  
citations

257101

24  
h-index

197535

49  
g-index

126  
all docs

126  
docs citations

126  
times ranked

3762  
citing authors

#	ARTICLE	IF	CITATIONS
1	Deformable Generator Networks: Unsupervised Disentanglement of Appearance and Geometry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44, 1162-1179.	9.7	5
2	Cascaded Parsing of Human-Object Interaction Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44, 2827-2840.	9.7	57
3	Synthesizing Diverse and Physically Stable Grasps With Arbitrary Hand Structures Using Differentiable Force Closure Estimator. IEEE Robotics and Automation Letters, 2022, 7, 470-477.	3.3	14
4	CX-ToM: Counterfactual explanations with theory-of-mind for enhancing human trust in image recognition models. IScience, 2022, 25, 103581.	1.9	19
5	Show Me What You Can Do: Capability Calibration on Reachable Workspace for Human-Robot Collaboration. IEEE Robotics and Automation Letters, 2022, 7, 2644-2651.	3.3	0
6	Understanding Physical Effects for Effective Tool-Use. IEEE Robotics and Automation Letters, 2022, 7, 9469-9476.	3.3	4
7	In situ bidirectional human-robot value alignment. Science Robotics, 2022, 7, .	9.9	27
8	Interpretable CNNs for Object Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 3416-3431.	9.7	25
9	A Generalized Earley Parser for Human Activity Parsing and Prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 2538-2554.	9.7	16
10	Learning Energy-Based Spatial-Temporal Generative ConvNets for Dynamic Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 516-531.	9.7	18
11	A unified psychological space for human perception of physical and social events. Cognitive Psychology, 2021, 128, 101398.	0.9	3
12	Hierarchical Human Semantic Parsing with Comprehensive Part-Relation Modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, PP, 1-1.	9.7	26
13	Extraction of an Explanatory Graph to Interpret a CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 3863-3877.	9.7	15
14	Reconstructing Interactive 3D Scenes by Panoptic Mapping and CAD Model Alignments. , 2021, , .		16
15	Mining Interpretable AOG Representations From Convolutional Networks via Active Question Answering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 3949-3963.	9.7	6
16	ACRE: Abstract Causal REasoning Beyond Covariation. , 2021, , .		8
17	Abstract Spatial-Temporal Reasoning via Probabilistic Abduction and Execution. , 2021, , .		12
18	Generative PointNet: Deep Energy-Based Learning on Unordered Point Sets for 3D Generation, Reconstruction and Classification. , 2021, , .		24

#	ARTICLE	IF	CITATIONS
19	Efficient Task Planning for Mobile Manipulation: a Virtual Kinematic Chain Perspective. , 2021, , .		10
20	Cooperative Training of Descriptor and Generator Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42, 27-45.	9.7	39
21	Representation Learning: A Statistical Perspective. Annual Review of Statistics and Its Application, 2020, 7, 303-335.	4.1	8
22	Joint Inference of States, Robot Knowledge, and Human (False-)Beliefs. , 2020, , .		11
23	Inducing Hierarchical Compositional Model by Sparsifying Generator Network. , 2020, , .		3
24	Joint Training of Variational Auto-Encoder and Latent Energy-Based Model. , 2020, , .		9
25	Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Humanlike Common Sense. Engineering, 2020, 6, 310-345.	3.2	56
26	A Competence-Aware Curriculum for Visual Concepts Learning via Question Answering. Lecture Notes in Computer Science, 2020, , 141-157.	1.0	10
27	LEMMA: A Multi-view Dataset for Learning Multi-agent Multi-task Activities. Lecture Notes in Computer Science, 2020, , 767-786.	1.0	13
28	Human-Robot Interaction in a Shared Augmented Reality Workspace. , 2020, , .		11
29	Graph-based Hierarchical Knowledge Representation for Robot Task Transfer from Virtual to Physical World. , 2020, , .		7
30	Understanding Human Gaze Communication by Spatio-Temporal Graph Reasoning. , 2019, , .		70
31	Holistic++ Scene Understanding: Single-View 3D Holistic Scene Parsing and Human Pose Estimation With Human-Object Interaction and Physical Commonsense. , 2019, , .		60
32	RAVEN: A Dataset for Relational and Analogical Visual REasoning. , 2019, , .		59
33	Reasoning Visual Dialogs With Structural and Partial Observations. , 2019, , .		78
34	Divergence Triangle for Joint Training of Generator Model, Energy-Based Model, and Inferential Model. , 2019, , .		18
35	Learning Dynamic Generator Model by Alternating Back-Propagation through Time. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 5498-5507.	3.6	17
36	A tale of two explanations: Enhancing human trust by explaining robot behavior. Science Robotics, 2019, 4, .	9.9	71

#	ARTICLE	IF	CITATIONS
37	Visual interpretability for deep learning: a survey. <i>Frontiers of Information Technology and Electronic Engineering</i> , 2018, 19, 27-39.	1.5	534
38	Attribute And-Or Grammar for Joint Parsing of Human Pose, Parts and Attributes. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2018, 40, 1555-1569.	9.7	42
39	Learning and Inferring "Dark Matter" and Predicting Human Intents and Trajectories in Videos. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2018, 40, 1639-1652.	9.7	37
40	Perception of Human Interaction Based on Motion Trajectories: From Aerial Videos to Decontextualized Animations. <i>Topics in Cognitive Science</i> , 2018, 10, 225-241.	1.1	14
41	Human-Centric Indoor Scene Synthesis Using Stochastic Grammar. , 2018, , .		81
42	Attentive Fashion Grammar Network for Fashion Landmark Detection and Clothing Category Classification. , 2018, , .		166
43	Learning Generative ConvNets via Multi-grid Modeling and Sampling. , 2018, , .		19
44	Learning Descriptor Networks for 3D Shape Synthesis and Analysis. , 2018, , .		63
45	Interpretable Convolutional Neural Networks. , 2018, , .		376
46	Unsupervised Learning of Hierarchical Models for Hand-Object Interactions. , 2018, , .		6
47	Interactive Robot Knowledge Patching Using Augmented Reality. , 2018, , .		47
48	Configurable 3D Scene Synthesis and 2D Image Rendering with Per-pixel Ground Truth Using Stochastic Grammars. <i>International Journal of Computer Vision</i> , 2018, 126, 920-941.	10.9	50
49	Learning Human-Object Interactions by Graph Parsing Neural Networks. <i>Lecture Notes in Computer Science</i> , 2018, , 407-423.	1.0	243
50	Sparse and deep generalizations of the FRAME model. <i>Annals of Mathematical Sciences and Applications</i> , 2018, 3, 211-254.	0.2	3
51	Modeling 4D Human-Object Interactions for Joint Event Segmentation, Recognition, and Object Localization. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2017, 39, 1165-1179.	9.7	49
52	A glove-based system for studying hand-object manipulation via joint pose and force sensing. , 2017, , .		29
53	Learning social affordance grammar from videos: Transferring human interactions to human-robot interactions. , 2017, , .		27
54	Synthesizing Dynamic Patterns by Spatial-Temporal Generative ConvNet. , 2017, , .		36

#	ARTICLE	IF	CITATIONS
55	Mining Object Parts from CNNs via Active Question-Answering. , 2017, , .		13
56	Feeling the force: Integrating force and pose for fluent discovery through imitation learning to open medicine bottles. , 2017, , .		33
57	Predicting Human Activities Using Stochastic Grammar. , 2017, , .		51
58	Multi-view People Tracking via Hierarchical Trajectory Composition. , 2016, , .		91
59	Inferring human intent from video by sampling hierarchical plans. , 2016, , .		14
60	Recognizing Car Fluents from Video. , 2016, , .		10
61	Inducing wavelets into random fields via generative boosting. Applied and Computational Harmonic Analysis, 2016, 41, 4-25.	1.1	10
62	Mining And-Or Graphs for Graph Matching and Object Discovery. , 2015, , .		9
63	Learning Sparse FRAME Models for Natural Image Patterns. International Journal of Computer Vision, 2015, 114, 91-112.	10.9	29
64	Understanding tools: Task-oriented object modeling, learning and recognition. , 2015, , .		84
65	Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face. , 2015, , .		33
66	Video Primal Sketch: A Unified Middle-Level Representation for Video. Journal of Mathematical Imaging and Vision, 2015, 53, 151-170.	0.8	5
67	Visual Persuasion: Inferring Communicative Intents of Images. , 2014, , .		66
68	Online Object Tracking, Learning, and Parsing with And-Or Graphs. , 2014, , .		43
69	Learning Inhomogeneous FRAME Models for Object Patterns. , 2014, , .		3
70	Single-View 3D Scene Parsing by Attributed Grammar. , 2014, , .		27
71	Learning AND-OR Templates for Object Recognition and Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35, 2189-2205.	9.7	84
72	Learning and parsing video events with goal and intent prediction. Computer Vision and Image Understanding, 2013, 117, 1369-1383.	3.0	21

#	ARTICLE	IF	CITATIONS
73	Modeling Occlusion by Discriminative AND-OR Structures. , 2013, , .		19
74	Integrating Grammar and Segmentation for Human Pose Estimation. , 2013, , .		27
75	Cosegmentation and Cosketch by Unsupervised Learning. , 2013, , .		43
76	Video Stylization: Painterly Rendering and Optimization With Content Extraction. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23, 577-590.	5.6	8
77	Learning Near-Optimal Cost-Sensitive Decision Policy for Object Detection. , 2013, , .		11
78	Scene Parsing by Integrating Function, Geometry and Appearance Models. , 2013, , .		101
79	Inferring &#x0022;Dark Matter&#x0022; and &#x0022;Dark Energy&#x0022; from Videos. , 2013, , .		18
80	Learning reconfigurable scene representation by tangram model. , 2012, , .		7
81	Reconfigurable templates for robust vehicle detection and classification. , 2012, , .		3
82	Unsupervised learning of stochastic AND-OR templates for object modeling. , 2011, , .		4
83	A Numerical Study of the Bottom-Up and Top-Down Inference Processes in And-Or Graphs. International Journal of Computer Vision, 2011, 93, 226-252.	10.9	53
84	Inferring social roles in long timespan video sequence. , 2011, , .		1
85	Unsupervised learning of event AND-OR grammar and semantics from video. , 2011, , .		60
86	Parsing video events with goal inference and intent prediction. , 2011, , .		95
87	Learning Active Basis Model for Object Detection and Recognition. International Journal of Computer Vision, 2010, 90, 198-235.	10.9	123
88	Discovering scene categories by information projection and cluster sampling. , 2010, , .		12
89	Learning a probabilistic model mixing 3D and 2D primitives for view invariant object recognition. , 2010, , .		20
90	Flow mosaicking: Real-time pedestrian counting without scene-specific learning. , 2009, , .		45

#	ARTICLE	IF	CITATIONS
91	Learning mixed templates for object recognition. , 2009, , .		5
92	Layered graph matching by composite cluster sampling with collaborative and competitive interactions. , 2009, , .		0
93	Flow mosaicking: Real-time pedestrian counting without scene-specific learning. , 2009, , .		0
94	Learning mixed templates for object recognition. , 2009, , .		0
95	Program chairs' introduction to the first international workshop on stochastic image grammars (SIG-09) in conjunction with IEEE CVPR 2009. , 2009, , .		0
96	Perceptual Scale-Space and Its Applications. International Journal of Computer Vision, 2008, 80, 143-165.	10.9	16
97	CCPR 2008 Keynote Speech 3 and Keynote Speech 4. , 2008, , .		0
98	Learning a scene contextual model for tracking and abnormality detection. , 2008, , .		12
99	An integrated background model for video surveillance based on primal sketch and 3D scene geometry. , 2008, , .		19
100	Generalizing Swendsen-Wang for Image Analysis. Journal of Computational and Graphical Statistics, 2007, 16, 877-900.	0.9	20
101	Statistical Principles in Image Modeling. Technometrics, 2007, 49, 249-261.	1.3	3
102	Mapping Natural Image Patches by Explicit and Implicit Manifolds. , 2007, , .		14
103	Layered Graph Match with Graph Editing. , 2007, , .		17
104	An Empirical Study of Object Category Recognition: Sequential Testing with Generalized Samples. , 2007, , .		21
105	Deformable Template As Active Basis. , 2007, , .		41
106	Compositional Boosting for Computing Hierarchical Image Structures. , 2007, , .		29
107	Parsing Images into Regions, Curves, and Curve Groups. International Journal of Computer Vision, 2006, 69, 223-249.	10.9	33
108	First Hitting Time Analysis of the Independence Metropolis Sampler. Journal of Theoretical Probability, 2006, 19, 235-261.	0.4	3

#	ARTICLE	IF	CITATIONS
109	A Stochastic Grammar of Images. Foundations and Trends in Computer Graphics and Vision, 2006, 2, 259-362.	2.8	273
110	Perceptual scale space and its applications. , 2005, , .		7
111	What are Textons?. International Journal of Computer Vision, 2005, 62, 121-143.	10.9	8
112	What are Textons?. International Journal of Computer Vision, 2005, 62, 121-143.	10.9	162
113	Image Parsing: Unifying Segmentation, Detection, and Recognition. International Journal of Computer Vision, 2005, 63, 113-140.	10.9	344
114	Incorporating visual knowledge representation in stereo reconstruction. , 2005, , .		5
115	Bottom-up/top-down image parsing by attribute graph grammar. , 2005, , .		40
116	Analysis and synthesis of textured motion: particles and waves. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26, 1348-1363.	9.7	28
117	Modeling Visual Patterns by Integrating Descriptive and Generative Methods. International Journal of Computer Vision, 2003, 53, 5-29.	10.9	60
118	Statistical modeling and conceptualization of visual patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25, 691-712.	9.7	81
119	Bayesian reconstruction of 3D shapes and scenes from a single image. , 2003, , .		29
120	Towards a mathematical theory of primal sketch and sketchability. , 2003, , .		32
121	A Generative Method for Textured Motion: Analysis and Synthesis. Lecture Notes in Computer Science, 2002, , 583-598.	1.0	29
122	Multigrid and multi-level swendsen-wang cuts for hierarchic graph partition. , 0, , .		13
123	Automatic Single View Building Reconstruction by Integrating Segmentation. , 0, , .		0
124	A High Resolution Grammatical Model for Face Representation and Sketching. , 0, , .		17
125	Cloth Representation by Shape from Shading with Shading Primitives. , 0, , .		1
126	Patching interpretable Andâ€œOrâ€œGraph knowledge representation using augmented reality. Applied AI Letters, 0, , e43.	1.4	2