

LIST OF DATASETS, SOFTWARE, LIBRARIES, AND MODELS

Table 1. List of datasets. Part I.

Name	Short Description	Link	Source
ACE	50 users exploring 5 different AR scences	https://cs.gmu.edu/~sqchen/open-access/ACE-Dataset.tgz	[62]
BigHand2.2M	Hand pose dataset	https://sites.google.com/site/qiyeincv/home/bibtex_cvpr2017	[67]
CASSIE Dataset	VR sketch data	https://gitlab.inria.fr/D3/cassie-data	[66]
Cityscape	Street scenes from 50 different cities	https://www.cityscapes-dataset.com/	[13]
CMU Graphics Lab Motion Capture Database	49 gaits obtained from subjects walking with different styles	http://mocap.cs.cmu.edu/	
CMU Panoptic Dataset	65 sequences and 1.5 millions of 3D skeletons	http://domedb.perception.cs.cmu.edu/	[25]
DeepFashion	Large-scale clothes database including annotations of clothing items and cross-pose/cross-domain image pairs	http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html	[34]
DGaze dataset	Gaze data in dynamic virtual indoor and outdoor scenes	http://zhiminghu.net/DGaze	[22]
Director's Cut	Includes the directional cues and plot points as well as the scan-paths of the test subjects watching films in VR	https://v-sense.scss.tcd.ie/?p=2477	[27]
DISFA	Spontaneous facial action intensity database	http://www.engr.du.edu/mmahoor/DISFA.htm	[38]
DIV2K	Diverse 2K resolution high quality images with a large diversity of contents	https://data.vision.ee.ethz.ch/cvl/DIV2K/	[1]
EgoCap	100.000 egocentric images of eight people in different clothing	https://vcai.mpi-inf.mpg.de/projects/EgoCap/	[48]
EgoVIP	Egocentric visual-inertial 3D human pose dataset	https://sites.google.com/site/youngwooncha/egovip	[8]
EHTaskDataset	Eye and head movements of 30 participants performing four tasks, i.e. Free viewing, Visual search, Saliency, and Track, in 15 360-degree VR videos	http://zhiminghu.net/EHTask	[21]
Enron Mobile Email Dataset	Sentences written by Enron employees on BlackBerry mobile devices	http://www.keithv.com/software/enronmobile/	[60]
Extended Cohn-Kanade Dataset (CK+)	Dataset for action unit and emotion-specified emotion	https://sites.pitt.edu/~emotion/ck-spread.htm	[35]
FERG-DB	2D images of stylized characters with annotated facial expressions	http://grail.cs.washington.edu/projects/deepexpr/ferg-2d-db.html	[?]
GrabAR1	Oaired images of hand and objects	<i>link not found</i>	[57]
GTSB	German traffic sign detection benchmark, including 900 images from three categories	https://benchmark.ini.rub.de/gtsdb_news.html	[20]
GTSRB	German traffic sign multi-category classification benchmark	https://benchmark.ini.rub.de/gtsrb_news.html	[55]
Human 3.6M	3.6 million human poses and corresponding images of 11 professional actors and 17 scenarios	http://vision.imar.ro/human3.6m/description.php	[23]
IISc Video Discomfort Dastate	videos and discomfort scores	https://github.com/rajiviisc/Video-Discomfort	[5]

Table 2. List of datasets. Part II.

Name	Short Description	Link	Source
ImageNet	Image database	https://www.image-net.org/challenges/LSVRC/	[51]
JAFPE	Japanese female facial expression dataset	https://zenodo.org/record/3451524	[36]
KITTI	Traffic scenarios	https://www.cvlibs.net/datasets/kitti/	[18]
Laval Indoor HDR Dataset	2100+ high resolution indoor panoramas	http://indoor.hdrdb.com/	[16]
Microsoft COCO: Common Objects in Context	Photos of 91 object types	https://arxiv.org/abs/1405.0312	[31]
MPI Emotional Body Expressions Database for Narrative Scenarios	Emotional body expressions	http://figshare.com/articles/MPI_EMB_M_Database_Mocap_Files/1220428	[61]
MPI-INF-3DHP	3D human body pose estimation dataset consisting of both constrained indoor and complex outdoor scenes	https://vcai.mpi-inf.mpg.de/3dhp-dataset/	[40]
MSRA14	Hand tracking dataset	https://jimmysuen.github.io/	[46]
MSRA15	Hand gesture dataset	https://jimmysuen.github.io/	[56]
PanoContext	Panorama dataset	https://panocontext.cs.princeton.edu/	[68]
People Snapshot Database	3D body models and texture of arbitrary people from a single, monocular video in which a person is moving	https://graphics.tu-bs.de/people-snapshot	[2]
Places2	Scene photographs of a diverse list of the types of environments	http://places2.csail.mit.edu/	[70]
Public-AR-Booksearch	Images of book spines in different size and various conditions	https://github.com/M-Schrapel/Public-AR-Booksearch	[52]
Stanford 2D-3D Semantics Dataset	Provides a variety of mutually registered modalities from 2D, 2.5D and 3D domains, with instance-level semantic and geometric annotations	http://buildingparser.stanford.edu/dataset.html	[3]
SUNCG	Synthetic 3D scenes	https://sscnet.cs.princeton.edu	[54]
The Million Song Dataset	Collection of audio features and metadata for a million contemporary popular music tracks	https://github.com/tbertinmahieux/MSongsDB	[6]
UEC FOOD 100	Food photos	http://foodcam.mobi/dataset100.html	[37]
UIBFEEED	Virtual facial expressions	http://ugivia.uib.es/uibvfed/	[?]
UNOC dataset	Large-scale motion capture dataset with body and finger motions	https://github.com/facebookresearch/UNOC	[44]
VR-EyeTracking	Eye tracking data of videos captured in dynamic scenes, each video is viewed by at least 31 subjects	https://github.com/xuyanyu-shh/VR-EyeTracking	[65]
VRSA	Image and video database	https://ivylabdb.kaist.ac.kr/	[26]
XR-Ego-Pose	Photorealistic egocentric camera images in a variety of indoor and outdoor space	https://github.com/facebookresearch/xR-EgoPose	[58]
-	LDR environment maps	http://www.jflalonde.ca/projects/deepIndoorLight	[17]

Table 3. List of datasets. Part III.

Name	Short Description	Link	Source
-	Colored 3D scans/Collection of points with 3D coordinates and RGB color values	http://buildingparser.stanford.edu/dataset.html	[4]
-	Stereoscopic 3D videos and their sickness ratings		[43]
-	Speech and corresponding gestures in a 3D human pose format	<i>no link found</i>	[47]
-	Visual-inertial input dataset for SLAM applications	https://doi.org/10.5281/zenodo.5018311	
-	Various datasets for viewport prediction	https://gitlab.com/miguelfromeror/head-motion-prediction/tree/master	[50]
-	Dataset for improving humans' ability to interpret deictic gestures in VR	https://github.com/interactionlab/Deictic-Pointing-in-VR	[39]
-	Human body motion reconstructing using only eyeglasses-mounted cameras and few body-worn inertial sensors	https://sites.google.com/site/youngwooncha/egovip	[9]
-	Exploring user behaviors in spherical video streaming	https://wuchlei-thu.github.io	[64]

Table 4. List of software toolkits and libraries. Part I.

Name	Short Description	Link	Paper
CERT: The Computer Expression Recognition Toolbox	"Fully automated facial expression recognition that operates in real-time"	https://inc.ucsd.edu/mplab/users/marini/Projects/CERT.htm	[32]
COVAREP	Repository for speech processing algorithms	http://covarep.github.io/covarep	[14]
Covert Embodied Choice	unity code for VR experimental setup	https://github.com/onejgordon/cec_vr	[19]
Daz-3D Studio	Creation of 3D scenes and characters	https://www.daz3d.com/	-
FAtiMA Toolkit	"Collection of tools/assets designed for the creation of characters and robots with social and emotional intelligence."	https://fatima-toolkit.eu/	-
Googles ARCore platform	"With ARCore, build new augmented reality experiences that seamlessly blend the digital and physical worlds. Transform the way people play, shop, learn, create, and experience the world together—at Google scale"	https://developers.google.com/ar	-
Google's Dialogflow service for dialogue manager	"Lifelike conversational AI with state-of-the-art virtual agents. Available in two editions: Dialogflow CX (advanced), Dialogflow ES (standard)"	https://cloud.google.com/dialogflow	-
HeMoG	gravitational white-box model for head motion estimation in 360 videos	https://gitlab.com/miguelfromeror/hemog	[49]
HRV Python library	"Heart Reate Variability analysis"	https://pypi.org/project/hrv-analysis/	-
Keras	"Keras is an API designed for human beings, not machines. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear & actionable error messages. It also has extensive documentation and developer guides."	https://keras.io/	-
Learning Gain Prediction	contains code and featurized data	https://github.com/LeonDong1993/learning-gain-prediction	[41]
LIPSYNC	Lip-syncing and facial animation tool for Unity	https://lipsync.rogodigital.com/	-
Mixamo	Animation tool for 3D character animation	https://www.mixamo.com/	-
OpenPose	"Real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints (in total 135 keypoints) on single images"	https://github.com/CMU-Perceptual-Computing-Lab/openpose	[7]

Table 5. List of software toolkits and libraries. PART II.

Name	Short Description	Link	Paper
OpenRDW	Provides APIs to access the attributes of scenes, to customize the RDW controllers, to simulate and visualize the navigation process, to export multiple formats of the results, and to evaluate RDW techniques	https://github.com/yaoling1997/OpenRDW	[30]
Panoptic-DeepLab	Image segmentation library	https://github.com/bowenc0221/panoptic-deeplab	[10]
PhysioNet	"The Research Resource for Complex Physiologic Signals"	https://physionet.org/	-
Poly Haven	3D asset library	https://hdrihaven.com/	-
PyTorch	"An open source machine learning framework that accelerates the path from research prototyping to production deployment."	https://pytorch.org/	-
ResonanceAudio	"Resonance Audio is a multi-platform spatial audio SDK, delivering high fidelity at scale. This powerful spatial audio technology is critical to realistic experiences for AR, VR, gaming, and video."	https://resonance-audio.github.io/resonance-audio/	-
Scikit-learn	"Simple and efficient tools for predictive data analysis Accessible to everybody, and reusable in various contexts Built on NumPy, SciPy, and matplotlib Open source, commercially usable - BSD license"	https://scikit-learn.org/stable/	-
TensorFlow	"Create production-grade machine learning models with TensorFlow"	https://www.tensorflow.org/	-
Shark library	"Shark is a fast, modular, feature-rich open-source C++ machine learning library. It provides methods for linear and nonlinear optimization, kernel-based learning algorithms, neural networks, and various other machine learning techniques. It serves as a powerful toolbox for real world applications as well as for research. Shark works on Windows, MacOS X, and Linux. It comes with extensive documentation. Shark is licensed under the GNU Lesser General Public License."	https://www.shark-ml.org/	-
Seurat	system for image-based scene simplification for VR	https://github.com/googlevr/seurat	[28]
SimSensei	Virtual interviewer for healthcare decision support	http://simsensei.ict.usc.edu/	[15]
VGG Image Annotator	Image annotator tool	https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html	-
Virtual Human Toolkit	Toolkit for the creation of virtual human conversational characters	https://vhtoolkit.ict.usc.edu/	-

Table 6. List of ML models and neural networks.

Name	Short Description	Link	Paper
ARShadowGAN	Model for creating virtual shadows	https://github.com/ldq9526/ARShadowGAN	[33]
BodyNet	Volumetric inference of 3D human body shapes	http://www.di.ens.fr/willow/research/bodynet/	[59]
Convolutional-Pose-Machines	Model for articulated pose estimation	https://github.com/CMU-Perceptual-Computing-Lab/convolutional-pose-machines-release	[63]
CUT	Contrastive unpaired translation for image-to-image translation	https://github.com/taesungp/contrastive-unpaired-translation	[45]
CycleGAN	Image-to-image translation without input-output pairs	https://github.com/junyanz/CycleGAN	[71]
EEGModels	A Collection of Convolutional Neural Network (CNN) models for EEG signal processing and classification, written in Keras and Tensorflow.	https://github.com/vlawhern/arl-eegmodels	[29]
ICNet	Model that creates segmentation masks for every pixel in an image	https://github.com/hellochick/ICNet-tensorflow	[69]
Pix2Pix	Image-to-image translation with conditional adversarial networks	https://github.com/phillipi/pix2pix	[24]
SiCloPe	Silhouette-based representation for modeling clothed human bodies	https://vgl.ict.usc.edu/Research/SiCloPe/	[42]
StarGAN	Image-to-image translations for multiple domains	https://github.com/yunjey/StarGAN	[11]
StarGAN v2	Image-to-image translations for multiple domains	https://github.com/clovaai/stargan-v2	[12]
-	Neural network for predicting avatar movements in VR	https://github.com/david-halbhuber/motionprediction	[53]

REFERENCES

- [1] Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [2] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. 2018. Video Based Reconstruction of 3D People Models. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8387–8397. <https://doi.org/10.1109/CVPR.2018.00875>
- [3] I. Armeni, A. Sax, A. R. Zamir, and S. Savarese. 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv e-prints* (Feb. 2017). arXiv:1702.01105 [cs.CV]
- [4] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 2016. 3D Semantic Parsing of Large-Scale Indoor Spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1534–1543. <https://doi.org/10.1109/CVPR.2016.170>
- [5] Suprith Balasubramanian and Rajiv Soundararajan. 2019. Prediction of Discomfort due to Egomotion in Immersive Videos for Virtual Reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.
- [6] Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. 2011. The million song dataset. (2011).
- [7] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).
- [8] Young-Woon Cha, Husam Shaik, Qian Zhang, Fan Feng, Andrei State, Adrian Ilie, and Henry Fuchs. 2021. Mobile. Egocentric Human Body Motion Reconstruction Using Only Eyeglasses-mounted Cameras and a Few Body-worn Inertial Sensors. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 616–625. <https://doi.org/10.1109/VR50410.2021.00087>
- [9] Young-Woon Cha, Husam Shaik, Qian Zhang, Fan Feng, Andrei State, Adrian Ilie, and Henry Fuchs. 2021. Mobile. Egocentric Human Body Motion Reconstruction Using Only Eyeglasses-mounted Cameras and a Few Body-worn Inertial Sensors. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 616–625.

- [10] Bowen Cheng, Maxwell D Collins, Yukun Zhu, Ting Liu, Thomas S Huang, Hartwig Adam, and Liang-Chieh Chen. 2020. Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation. In *CVPR*.
- [11] Yunje Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2018. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [12] Yunje Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. 2020. StarGAN v2: Diverse Image Synthesis for Multiple Domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [13] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. *CoRR* abs/1604.01685 (2016). arXiv:1604.01685 <http://arxiv.org/abs/1604.01685>
- [14] Gilles Degottex, John Kane, Thomas Drugman, Tuomo Raitio, and Stefan Scherer. 2014. COVAREP – A collaborative voice analysis repository for speech technologies. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 960–964. <https://doi.org/10.1109/ICASSP.2014.6853739>
- [15] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems (Paris, France) (AAMAS '14)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1061–1068.
- [16] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to Predict Indoor Illumination from a Single Image. *CoRR* abs/1704.00090 (2017). arXiv:1704.00090 <http://arxiv.org/abs/1704.00090>
- [17] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to Predict Indoor Illumination from a Single Image. *CoRR* abs/1704.00090 (2017). arXiv:1704.00090 <http://arxiv.org/abs/1704.00090>
- [18] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* 32, 11 (2013), 1231–1237.
- [19] Jeremy Raboff Gordon, Max T. Curran, John Chuang, and Coye Cheshire. 2021. Covert Embodied Choice: Decision-Making and the Limits of Privacy Under Biometric Surveillance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 551, 12 pages. <https://doi.org/10.1145/3411764.3445309>
- [20] Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel. 2013. Detection of Traffic Signs in Real-World Images: The German Traffic Sign Detection Benchmark. In *International Joint Conference on Neural Networks*.
- [21] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. EHTask: Recognizing User Tasks from Eye and Head Movements in Immersive Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. <https://doi.org/10.1109/TVCG.2021.3138902>
- [22] Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. 2020. DGaze: CNN-Based Gaze Prediction in Dynamic Scenes. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1902–1911.
- [23] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. 2014. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (jul 2014), 1325–1339.
- [24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *CoRR* abs/1611.07004 (2016). arXiv:1611.07004 <http://arxiv.org/abs/1611.07004>
- [25] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Scott Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. 2017. Panoptic Studio: A Massively Multiview System for Social Interaction Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).
- [26] Hak Gu Kim, Heoun-Taek Lim, Sangmin Lee, and Yong Man Ro. 2019. VRSA Net: VR Sickness Assessment Considering Exceptional Motion for 360° VR Video. *IEEE Transactions on Image Processing* 28, 4 (2019), 1646–1660. <https://doi.org/10.1109/TIP.2018.2880509>
- [27] Sebastian Knorr, Cagri Ozcinar, Colm O Fearghail, and Aljosa Smolic. 2018. Director's Cut - A Combined Dataset for Visual Attention Analysis in Cinematic VR Content. In *The 15th ACM SIGGRAPH European Conference on Visual Media Production*.
- [28] Puneet Lall, Silviu Borac, Dave Richardson, Matt Pharr, and Manfred Ernst. 2018. View-Region Optimized Image-Based Scene Simplification. *Proc. ACM Comput. Graph. Interact. Tech.* 1, 2, Article 26 (aug 2018), 22 pages. <https://doi.org/10.1145/3233311>
- [29] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering* 15, 5 (2018), 056013. <http://stacks.iop.org/1741-2552/15/i=5/a=056013>
- [30] Yi-Jun Li, Miao Wang, Frank Steinicke, and Qinping Zhao. 2021. OpenRDW: A Redirected Walking Library and Benchmark with Multi-User, Learning-based Functionalities and State-of-the-art Algorithms. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 21–30. <https://doi.org/10.1109/ISMAR52148.2021.00016>
- [31] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. <https://doi.org/10.48550/ARXIV.1405.0312>
- [32] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. 2011. The computer expression recognition toolbox (CERT). In *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*. 298–305. <https://doi.org/10.1109/FG.2011.5771414>
- [33] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhong Dong, and Chunxia Xiao. 2020. ARShadowGAN: Shadow Generative Adversarial Network for Augmented Reality in Single Light Scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

- [34] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [35] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. 2010. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. 94–101. <https://doi.org/10.1109/CVPRW.2010.5543262>
- [36] Michael J. Lyons. 2021. "Excavating AI" Re-excavated: Debunking a Fallacious Account of the JAFFE Dataset. *CoRR* abs/2107.13998 (2021). arXiv:2107.13998 <https://arxiv.org/abs/2107.13998>
- [37] Y. Matsuda, H. Hoashi, and K. Yanai. 2012. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*.
- [38] S. Mohammad Mavadati, Mohammad H. Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F. Cohn. 2013. DISFA: A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing* 4, 2 (2013), 151–160. <https://doi.org/10.1109/T-AFFC.2013.4>
- [39] Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelle, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020-04-25) (CHI '20)*. ACM, New York, NY, USA. <https://doi.org/10.1145/3313831.3376340>
- [40] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. 2017. Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision. In *3D Vision (3DV), 2017 Fifth International Conference on*. IEEE. <https://doi.org/10.1109/3dv.2017.00064>
- [41] Alec G. Moore, Ryan P. McMahan, Hailiang Dong, and Nicholas Ruozzi. 2020. Extracting Velocity-Based User-Tracking Features to Predict Learning Gains in a Virtual Reality Training Application. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 694–703. <https://doi.org/10.1109/ISMAR50242.2020.00099>
- [42] Ryota Natsume, Shunsuke Saito, Zeng Huang, Weikai Chen, Chongyang Ma, Hao Li, and Shigeo Morishima. 2019. SiCloPe: Silhouette-Based Clothed People. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [43] Nitish Padmanaban, Timon Ruban, Vincent Sitzmann, Anthony M. Norcia, and Gordon Wetzstein. 2018. Towards a Machine-Learning Approach for Sickness Prediction in 360° Stereoscopic Videos. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1594–1603. <https://doi.org/10.1109/TVCG.2018.2793560>
- [44] Mathias Parger, Chengcheng Tang, Yuanlu Xu, Christopher David Twigg, Lingling Tao, Yijing Li, Robert Wang, and Markus Steinberger. 2021. UNOC: Understanding Occlusion for Embodied Presence in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. <https://doi.org/10.1109/TVCG.2021.3085407>
- [45] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. 2020. Contrastive Learning for Unpaired Image-to-Image Translation. In *European Conference on Computer Vision*.
- [46] Chen Qian, Xiao Sun, Yichen Wei, Xiaoou Tang, and Jian Sun. 2014. Realtime and Robust Hand Tracking from Depth. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [47] Manuel Rebol, Christian Güti, and Krzysztof Pietroszek. 2021. Passing a Non-verbal Turing Test: Evaluating Gesture Animations Generated from Speech. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 573–581. <https://doi.org/10.1109/VR50410.2021.00082>
- [48] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. 2016. EgoCap: Egocentric Marker-Less Motion Capture with Two Fisheye Cameras. *ACM Trans. Graph.* 35, 6, Article 162 (nov 2016), 11 pages. <https://doi.org/10.1145/2980179.2980235>
- [49] Miguel Fabian Romero Rondon, Dario Zanca, Stefano Melacci, Marco Gori, and Lucile Sassatelli. 2021. HeMoG: A White-Box Model to Unveil the Connection between Saliency Information and Human Head Motion in Virtual Reality. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 10–18. <https://doi.org/10.1109/AIVR52153.2021.00012>
- [50] Miguel Fabián Romero Rondón, Lucile Sassatelli, Ramón Aparicio-Pardo, and Frédéric Precioso. 2022. TRACK: A New Method From a Re-Examination of Deep Architectures for Head Motion Prediction in 360 Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 9 (2022), 5681–5699. <https://doi.org/10.1109/TPAMI.2021.3070520>
- [51] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- [52] Maximilian Schrapel, Thilo Schulz, and Michael Rohs. 2020. Augmenting Public Bookcases to Support Book Sharing. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (Oldenburg, Germany) (MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. <https://doi.org/10.1145/3379503.3403542>
- [53] Valentin Schwind, David Halbhuber, Jakob Fehle, Jonathan Sasse, Andreas Pfaffelhuber, Christoph Tögel, Julian Dietz, and Niels Henze. 2020. The Effects of Full-Body Avatar Movement Predictions in Virtual Reality Using Neural Networks. In *26th ACM Symposium on Virtual Reality Software and Technology (Virtual Event, Canada) (VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 28, 11 pages. <https://doi.org/10.1145/3385956.3418941>
- [54] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. 2017. Semantic Scene Completion from a Single Depth Image. *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition* (2017).

- [55] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks* 0 (2012), –. <https://doi.org/10.1016/j.neunet.2012.02.016>
- [56] Xiao Sun, Yichen Wei, Shuang Liang, Xiaou Tang, and Jian Sun. 2015. Cascaded Hand Pose Regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [57] Xiao Tang, Xiaowei Hu, Chi-Wing Fu, and Daniel Cohen-Or. 2020. GrabAR: Occlusion-Aware Grabbing Virtual Objects in AR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 697–708. <https://doi.org/10.1145/3379337.3415835>
- [58] Denis Tome, Patrick Peluse, Lourdes Agapito, and Hernan Badino. 2019. xR-EgoPose: Egocentric 3D Human Pose from an HMD Camera. In *Proceedings of the IEEE International Conference on Computer Vision*. 7728–7738.
- [59] Gul Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. 2018. BodyNet: Volumetric Inference of 3D Human Body Shapes. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- [60] Keith Vertanen and Per Ola Kristensson. 2011. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (Stockholm, Sweden) (MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 295–298. <https://doi.org/10.1145/2037373.2037418>
- [61] Ekaterina Volkova, Stephan de la Rosa, Heinrich H. Bühlhoff, and Betty Mohler. 2014. The MPI Emotional Body Expressions Database for Narrative Scenarios. *PLOS ONE* 9, 12 (Dec. 2014), 1–28. <https://doi.org/10.1371/journal.pone.0113647> Publisher: Public Library of Science.
- [62] Na Wang, Haoliang Wang, Stefano Petrangeli, Viswanathan Swaminathan, Fei Li, and Songqing Chen. 2020. Towards Field-of-View Prediction for Augmented Reality Applications on Mobile Devices. In *Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems (Istanbul, Turkey) (MMVE '20)*. Association for Computing Machinery, New York, NY, USA, 13–18. <https://doi.org/10.1145/3386293.3397114>
- [63] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. 2016. Convolutional pose machines. In *CVPR*.
- [64] Chenglei Wu, Zhihao Tan, Zhi Wang, and Shiqiang Yang. 2017. A Dataset for Exploring User Behaviors in VR Spherical Video Streaming. In *Proceedings of the 8th ACM on Multimedia Systems Conference (Taipei, Taiwan) (MMSys'17)*. Association for Computing Machinery, New York, NY, USA, 193–198. <https://doi.org/10.1145/3083187.3083210>
- [65] Yanyu Xu, Yanbing Dong, Junru Wu, Zhengzhong Sun, Zhiru Shi, Jingyi Yu, and Shenghua Gao. 2018. Gaze Prediction in Dynamic 360° Immersive Videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5333–5342. <https://doi.org/10.1109/CVPR.2018.00559>
- [66] Emilie Yu, Rahul Arora, Tibor Stanko, J. Andreas Bærentzen, Karan Singh, and Adrien Bousseau. 2021. CASSIE: Curve and Surface Sketching in Immersive Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 190, 14 pages. <https://doi.org/10.1145/3411764.3445158>
- [67] Shanxin Yuan, Qi Ye, Bjorn Stenger, Siddhant Jain, and Tae-Kyun Kim. 2017. BigHand2.2M Benchmark: Hand Pose Dataset and State of the Art Analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [68] Yinda Zhang, Shuran Song, Ping Tan, and Jianxiong Xiao. 2014. PanoContext: A Whole-Room 3D Context Model for Panoramic Scene Understanding. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 668–686.
- [69] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. 2017. ICNet for Real-Time Semantic Segmentation on High-Resolution Images. *CoRR* abs/1704.08545 (2017). arXiv:1704.08545 <http://arxiv.org/abs/1704.08545>
- [70] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017). Publisher: IEEE.
- [71] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networkss. In *Computer Vision (ICCV), 2017 IEEE International Conference on*.