

fellowship from la Caixa Foundation (ID 100010434) under the fellowship code LCF/BQ/IN18/ 11660029. This research was also partially supported by Facebook Inc. The author would like to thank all the collaborators of this project and in particular Shruti Palaskar, Kenneth Joseph DeHaan, Deepti Ghadiyaram and Yannis Kalantidis for all the insightful discussions and help in the development of the project.

REFERENCES

[1] World Health Organization 2019. [n.d.]. *Deafness and hearing loss*. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>

[2] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan, and Ashwin Thangali. 2008. The american sign language lexicon video dataset. In *CVPRW'08*. IEEE, 1–8.

[3] Patrick Buehler, Andrew Zisserman, and Mark Everingham. 2009. Learning sign language by watching TV (using weakly aligned subtitles). In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2961–2968.

[4] Jan Bungeroth, Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, Andy Way, and Lynette van Zijl. 2008. The ATIS sign language corpus. *6th International Conference on Language Resources and Evaluation* (2008).

[5] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008* (2018).

[6] N Caselli, Z Sevcikova, A Cohen-Goldberg, and K Emmorey. 2016. ASL-Lex: A lexical database for ASL. *Behavior Research Methods* (2016).

[7] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros. 2018. Everybody dance now. *arXiv preprint arXiv:1808.07371* (2018).

[8] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *Conference on Computer Vision and Pattern Recognition*. 7784–7793.

[9] Reuben Conrad. 1979. *The deaf schoolchild: Language and cognitive function*. HarperCollins Publishers.

[10] Helen Cooper and Richard Bowden. 2009. Learning signs from subtitles: A weakly supervised approach to sign language recognition. In *CVPR*. IEEE, 2568–2574.

[11] Helen Cooper, Eng-Jon Ong, Nicolas Pugeault, and Richard Bowden. 2012. Sign language recognition using sub-units. *Journal of Machine Learning Research* 13, Jul (2012), 2205–2231.

[12] Stephen Cox, Michael Lincoln, Judy Tryggvason, Melanie Nakisa, Mark Wells, Marcus Tutt, and Sanja Abbott. 2002. Tessa, a system to aid communication with deaf people. In *5th International ACM conference on Assistive technologies*. ACM, 205–212.

[13] Rungpeng Cui, Hu Liu, and Changshui Zhang. 2017. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[14] Mary Jo Davidson. 2006. PAULA: A computer-based sign language tutor for hearing adults. In *Workshop on Teaching with Robots, Agents, and Natural Language Processing*.

[15] Philippe Dreuw, Thomas Deselaers, Daniel Keysers, and Hermann Ney. 2006. Modeling image variability in appearance-based gesture recognition. In *ECCV Workshop on Statistical Methods in Multi-Image and Video Processing*. 7–18.

[16] Adil Er-Rady, R Faizi, R Oulad Haj Thami, and H Housni. 2017. Automatic sign language recognition: A survey. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2017 International Conference on*. IEEE, 1–7.

[17] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus H Piater, and Hermann Ney. 2012. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus.. In *LREC*. 3785–3789.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

[19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[20] Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. 2015. Panoptic studio: A massively multi-view system for social motion capture. In *Proceedings of the IEEE International Conference on Computer Vision*. 3334–3342.

[21] Saba Joudaki, Dzulkifli bin Mohamad, Tanzila Saba, Amjad Rehman, Mz-nah Al-Rodhaan, and Abdullah Al-Dhelaan. 2014. Vision-Based Sign Language Classification: A Directional Review. *IETE Technical Review* 31, 5 (2014), 383–391. <https://doi.org/10.1080/02564602.2014.961576> arXiv:<https://doi.org/10.1080/02564602.2014.961576>

[22] Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *International Workshop on Intelligent Virtual Agents*. Springer, 113–126.

[23] Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2018. Neural Sign Language Translation based on Human Keypoint Estimation. *arXiv preprint*

arXiv:1811.11436 (2018).

[24] Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural Sign Language Translation Based on Human Keypoint Estimation. *Applied Sciences* 9, 13 (2019).

[25] Sang-Ki Ko, Jae Gi Son, and Hyedong Jung. 2018. Sign language recognition with recurrent neural network using human keypoint detection. In *Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems*. ACM, 326–328.

[26] Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding* 141 (2015), 108–125.

[27] Oscar Koller, Hermann Ney, and Richard Bowden. 2015. Deep learning of mouth shapes for sign language. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 85–91.

[28] Oscar Koller, Hermann Ney, and Richard Bowden. 2016. Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In *Conference on Computer Vision and Pattern Recognition*. 3793–3802.

[29] Oscar Koller, O Zargaran, Hermann Ney, and Richard Bowden. 2016. Deep sign: hybrid CNN-HMM for continuous sign language recognition. In *British Machine Vision Conference 2016*.

[30] Jolanta Lapiak. [n.d.]. American Sign Language Dictionary. <https://www.handspeak.com/word/>. Accessed: 2019-08-20.

[31] Jolanta Lapiak. [n.d.]. Gloss: transcription symbols. <https://www.handspeak.com/learn/index.php?id=3>. Accessed: 2019-08-20.

[32] Scott K Liddell et al. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.

[33] Alex M Martinez, Ronnie B Wilbur, Robin Shay, and Avinash C Kak. 2002. Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In *4th IEEE International Conference on Multimodal Interfaces, 2002*. IEEE, 167–172.

[34] Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 2012. Dicta-Sign—building a multilingual sign language corpus. In *5th LREC*. Istanbul.

[35] Eng-Jon Ong, Helen Cooper, Nicolas Pugeault, and Richard Bowden. 2012. Sign language recognition using sequential pattern trees. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2200–2207.

[36] Eng-Jon Ong, Helen Cooper, Nicolas Pugeault, and Richard Bowden. 2012. Sign Language Recognition using Sequential Pattern Trees. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. Providence, Rhode Island, USA. http://personal.ee.surrey.ac.uk/Personal/H.Cooper/research/papers/Ong_Sign_2012.pdf

[37] Manny Rayner, Pierrette Bouillon, Sarah Ebling, Johanna Gerlach, Irene Strasly, and Nikos Tsourakis. 2016. An open web platform for rule-based speech-to-sign translation. In *54th Annual Meeting of the Association for Computational Linguistics*, Vol. 2. 162–168.

[38] Rubén San-Segundo, Juan Manuel Montero, Javier Macías-Guarasa, R Córdoba, Javier Ferreiros, and José Manuel Pardo. 2008. Proposing a speech to gesture translation architecture for Spanish deaf people. *Journal of Visual Languages & Computing* 19, 5 (2008), 523–538.

[39] Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loïc Barrault, Lucia Specia, and Florian Metz. 2018. How2: a large-scale dataset for multimodal language understanding. *arXiv preprint arXiv:1811.00347* (2018).

[40] Adam Schembri, Jordan Fenlon, Ramas Rentelis, Sally Reynolds, and Kearsy Cormier. 2013. Building the British sign language corpus. *Language Documentation & Conservation* 7 (2013), 136–154.

[41] William C Stokoe Jr. 2005. Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of deaf studies and deaf education* 10, 1 (2005), 3–37.

[42] Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2018. Sign Language Production using Neural Machine Translation and Generative Adversarial Networks.. In *BMVC*. 304.

[43] Clayton Valli and Ceil Lucas. 2000. *Linguistics of American sign language: an introduction*. Gallaudet University Press.

[44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[45] Christian Vogler and Dimitris Metaxas. 1999. Parallel hidden markov models for american sign language recognition. In *7th IEEE International Conference on Computer Vision*, Vol. 1. IEEE, 116–122.

[46] David Wood, Heather Wood, Amanda Griffiths, and Ian Howarth. 1995. *Teaching and talking with deaf children*. Wiley.

[47] Hee-Deok Yang, Stan Sclaroff, and Seong-Whan Lee. 2009. Sign language spotting with a threshold model based on conditional random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 7 (2009), 1264–1277.

[48] Jan Zelinka, Jakub Kanis, and Petr Salajka. 2019. NN-Based Czech Sign Language Synthesis. In *International Conference on Speech and Computer*. Springer, 559–568.

[49] Jan Zelinka, Jakub Kanis, and Petr Salajka. 2019. NN-Based Czech Sign Language Synthesis. In *International Conference on Speech and Computer*. Springer, 559–568.