REFERENCES

[1] Qian Bao, Wu Liu, Jun Hong, Lingyu Duan, and Tao Mei. 2020. Pose-native Network Architecture Search for Multi-person Human Pose Estimation. In *ACM International Conference on Multimedia*. 592–600.

[2] Donald J Bemdt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *AAAI Workshop on Knowledge Discovery in Databases*. 359–370.

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

[4] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *IEEE Conference on Computer Vision and Pattern Recognition*. 7291–7299.

[5] Runpeng Cui, Zhong Cao, Weishen Pan, Changshui Zhang, and Jianqiang Wang. 2019. Deep Gesture Video Generation with Learning on Regions of Interest. *IEEE Transactions on Multimedia* 22, 10 (2019), 2551–2563.

[6] Mathieu De Coster, Mieke Van Herreweghe, and Joni Dambre. 2021. Isolated Sign Recognition from RGB Video Using Pose Flow and Self-attention. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3441–3450.

[7] JRW Glauert, Ralph Elliott, SJ Cox, Judy Tryggvason, and Mary Sheard. 2006. VANESSA – A System for Communication Between Deaf and Hearing People. *Technology and Disability* 18, 4 (2006), 207–216.

[8] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In *International Conference on Machine Learning*. 369–376.

[9] Dan Guo, Shengeng Tang, and Meng Wang. 2019. Connectionist temporal modeling of video and language: a joint model for translation and sign labeling. In *International Joint Conference on Artificial Intelligence*. 751–757.

[10] Dan Guo, Wengang Zhou, Houqiang Li, and Meng Wang. 2017. Online Early-late Fusion Based on Adaptive HMM for Sign Language Recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications* 14, 1 (2017), 1–18.

[11] Wencan Huang, Wenwen Pan, Zhou Zhao, and Qi Tian. 2021. Towards Fast and High-Quality Sign Language Production. In *ACM International Conference on Multimedia*. 3172–3181.

[12] Euijun Hwang, Jung-Ho Kim, and Jong-Cheol Park. 2021. Non-Autoregressive Sign Language Production with Gaussian Space. In *British Machine Vision Conference*.

[13] Kostas Karpouzis, George Caridakis, S-E Fotinea, and Eleni Efthimiou. 2007. Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture. *Computers & Education* 49, 1 (2007), 54–74.

[14] Dilek Kayahan and Tunga Güngör. 2019. A Hybrid Translation System from Turkish Spoken Language to Turkish Sign Language. In *International Symposium on INnovations in Intelligent SysTems and Applications*. 1–6.

[15] Diederik P Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*.

[16] Dimitris Kouremenos, Klimis S Ntalianis, Giorgos Siolas, and Andreas Stafylopatis. 2018. Statistical Machine Translation for Greek to Greek Sign Language Using Parallel Corpora Produced via Rule-Based Machine Translation. In *International Conference on Tools with Artificial Intelligence*. 28–42.

[17] Sven Kreiss, Lorenzo Bertoni, and Alexandre Alahi. 2019. Pifpaf: Composite Fields for Human Pose Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*. 11977–11986.

[18] Shyam Krishna and Janmesh Ukey. 2021. GAN Based Indian Sign Language Synthesis. In *Indian Conference on Vision, Graphics and Image Processing*. 1–8.

[19] Taro Miyazaki, Yusuke Morita, and Masanori Sano. 2020. Machine Translation from Spoken Language to Sign Language Using Pre-trained Language Model As Encoder. In *Workshop on the Representation and Processing of Sign Languages*. 139–144.

[20] Amit Moryossef, Kayo Yin, Graham Neubig, and Yoav Goldberg. 2021. Data Augmentation for Sign Language Gloss Translation. In *International Workshop on Automatic Translation for Signed and Spoken Languages*. 1–11.

[21] B Natarajan and R Elakkiya. 2022. Dynamic GAN for High-Quality Sign Language Video Generation from Skeletal Poses Using Generative Adversarial Networks. *Soft Computing* (2022).

[22] Achraf Othman and Mohamed Jemni. 2011. Statistical Sign Language Machine Translation: from English written text to American Sign Language Gloss. *International Journal of Computer Science Issues* 8, 5 (2011), 65–73.

[23] Gabriele Prato, Ella Charlaix, and Mehdi Rezagholizadeh. 2020. Fully Quantized Transformer for Machine Translation. In *Conference on Empirical Methods in Natural Language Processing*. 1–14.

[24] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning Transferable Visual Models from Natural Language Supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[25] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020. Adversarial Training for Multi-Channel Sign Language Production. In *British Machine Vision Conference*.

[26] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020. Progressive Transformers for End-to-end Sign Language Production. In *European Conference on Computer Vision*. 687–705.

[27] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2021. AnonySign: Novel Human Appearance Synthesis for Sign Language Video Anonymisation. In *IEEE International Conference on Automatic Face and Gesture Recognition*. 1–8.

[28] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2021. Continuous 3d Multi-channel Sign Language Production Via Progressive Transformers and Mixture Density Networks. *International Journal of Computer Vision* 129, 7 (2021), 2113–2135.

[29] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2021. Mixed SIGNals: Sign Language Production via a Mixture of Motion Primitives. In *IEEE International Conference on Computer Vision*. 1919–1929.

[30] Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. 2020. Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. *International Journal of Computer Vision* 128, 4 (2020), 891–908.

[31] Stephanie Stoll, Simon Hadfield, and Richard Bowden. 2020. SignSynth: Data-Driven Sign Language Video Generation. In *European Conference on Computer Vision*. 353–370.

[32] Weijie Su, Xizhou Zhu, Yue Cao, Bin Li, Lewei Lu, Furu Wei, and Jifeng Dai. 2019. VL-BERT: Pre-training of Generic Visual-Linguistic Representations. In *International Conference on Learning Representations*.

[33] Shengeng Tang, Dan Guo, Richang Hong, and Meng Wang. 2021. Graph-Based Multimodal Sequential Embedding for Sign Language Translation. *IEEE Transactions on Multimedia* (2021).

[34] Inigo Jauregi Unanue, Jacob Parnell, and Massimo Piccardi. 2021. BERTTune: Fine-Tuning Neural Machine Translation with BERTScore. In *Annual Meeting of the Association for Computational Linguistics*. 915–924.

[35] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[36] Neel Vasani, Pratik Autee, Samip Kalyani, and Ruhina Karani. 2020. Generation of Indian Sign Language by Sentence Processing and Generative Adversarial Networks. In *International Conference on Information Systems Security*. 1250–1255.

[37] 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 *Annual Conference on Neural Information Processing Systems*.

[38] Lucas Ventura, Amanda Duarte, and Xavier Giró-i Nieto. 2020. Can Everybody Sign Now? Exploring Sign Language Video Generation from 2d Poses. In *Sign Language Recognition, Translation & Production*.

[39] Carla Viegas, Mert Inan, Lorna Quandt, and Malihe Alikhani. 2022. Including Facial Expressions in Contextual Embeddings for Sign Language Generation. *arXiv preprint arXiv:2202.05383* (2022).

[40] Hanjie Wang, Xiujuan Chai, and Xilin Chen. 2019. A Novel Sign Language Recognition Framework Using Hierarchical Grassmann Covariance Matrix. *IEEE Transactions on Multimedia* 21, 11 (2019), 2806–2814.

[41] Yaohui Wang, Piotr Bilinski, Francois Bremond, and Antitza Dantcheva. 2020. G3AN: Disentangling Appearance and Motion for Video Generation. In *IEEE Conference on Computer Vision and Pattern Recognition*. 5264–5273.

[42] Qinkun Xiao, Minying Qin, and Yuting Yin. 2020. Skeleton-based Chinese Sign Language Recognition and Generation for Bidirectional Communication Between Deaf and Hearing People. *Neural Networks* 125 (2020), 41–55.

[43] Ceyuan Yang, Zhe Wang, Xinge Zhu, Chen Huang, Jianping Shi, and Dahua Lin. 2018. Pose Guided Human Video Generation. In *European Conference on Computer Vision*. 201–216.

[44] Jiacheng Yang, Mingxuan Wang, Hao Zhou, Chengqi Zhao, Weinan Zhang, Yong Yu, and Lei Li. 2020. Towards Making the Most of Bert in Neural Machine Translation. In *AAAI Conference on Artificial Intelligence*, Vol. 34. 9378–9385.

[45] Jan Zelinka and Jakub Kanis. 2020. Neural Sign Language Synthesis: Words Are Our Glosses. In *International Workshop on Applications of Computer Vision*. 3395–3403.

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

[47] Jiali Zeng, Shuangzhi Wu, Yongjing Yin, Yufan Jiang, and Mu Li. 2021. Recurrent Attention for Neural Machine Translation. In *Conference on Empirical Methods in Natural Language Processing*. 3216–3225.

[48] Ni Zeng, Yiqiang Chen, Yang Gu, Dongdong Liu, and Yunbing Xing. 2020. Highly Fluent Sign Language Synthesis Based on Variable Motion Frame Interpolation. In *IEEE International Conference on Systems, Man, and Cybernetics*. 1772–1777.

[49] Tianfu Zhang, He-Yan Huang, Chong Feng, and Longbing Cao. 2021. Enlivening Redundant Heads in Multi-head Self-attention for Machine Translation. In *Conference on Empirical Methods in Natural Language Processing*. 3238–3248.