

DeepLip: A Comprehensive Review on Deep Learning Based Lip Reading for Visual Speech Recognition

Nikhil Jagtap¹, Shreya Banarase², Mitali Kucheriya³, Vivek Kumavat⁴

¹⁻⁴Department of Computer Engineering,

ISBM College of Engineering, Nande, Pune, India

jagtapnikhil2411@gmail.com, shreyabanarase123@gmail.com,

mitalikucheriya@gmail.com, vivekkumawat1304@gmail.com

Abstract: Visual speech recognition, or lip reading, is one of the fastest growing domains in computer vision and speech processing. Lip reading has transitioned from rule-based models to robust, fully end-to-end trainable systems through deep learning. In this review paper, we systematically summarize the evolution of lip reading from early hand-crafted feature extraction to current architectures for lip reading research as methods such as 3D CNN, Residual Networks, and Transformer-based methods. We define the state of the art for lip reading in terms of datasets, algorithms, evaluation metrics, and implementation. We also discuss open questions, ethical issues, and potential future directions to further progress the field of visual speech perception.

Keywords: Lip reading, visual speech recognition, deep learning, convolutional neural network, trans-former, datasets

