Survey on Automatic Lip-Reading in the Era of Deep Learning - Supplementary Materials -

Adriana Fernandez-Lopez and Federico Sukno

Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, Spain.

In the supplementary materials, we provide the block diagrams of the most representative end-to-end deep learning systems up to 2017. For each figure, the left-most blocks describe the architecture configuration of each type of network. The remaining blocks share the same configuration, hence only the network type is indicated.

For the WLAS architecture presented by Chung et al. [16] we provide three separate figures. Figure S6 shows the block diagram of the whole system, while Figures S7 and S8 show the details of the Watch and Spell networks, respectively.

Email address: adriana.fernandez@upf.edu, federico.sukno@upf.edu (Adriana Fernandez-Lopez and Federico Sukno)

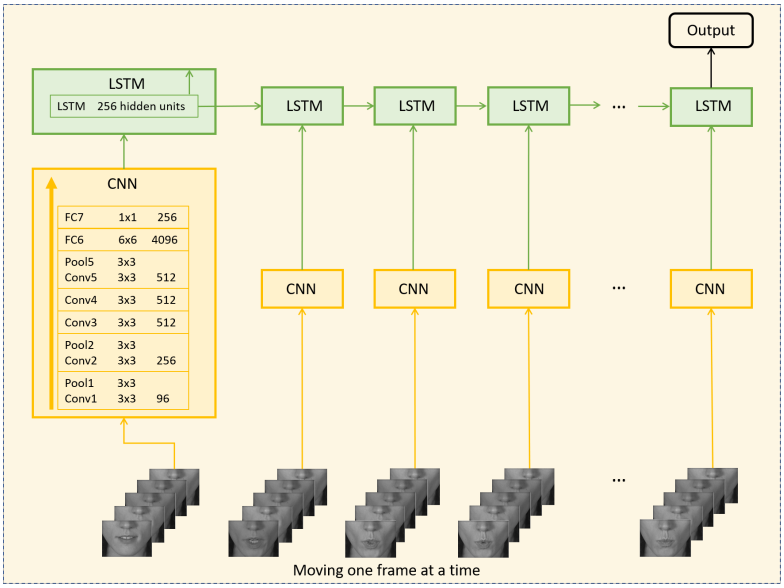

Figure S1: Architecture from Chung et al. [146] - Combination of SyncNet and LSTM networks

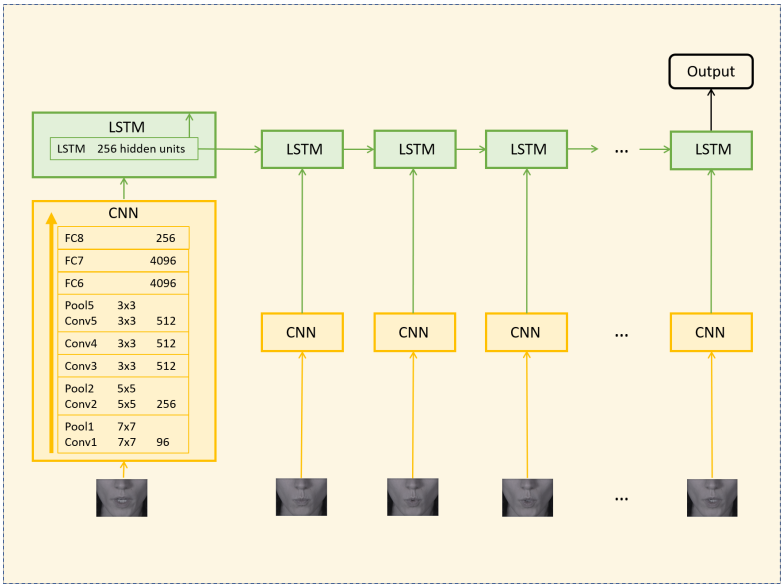

Figure S2: Architecture from Chung et al. [146] - Combination of VGG-M and LSTM networks

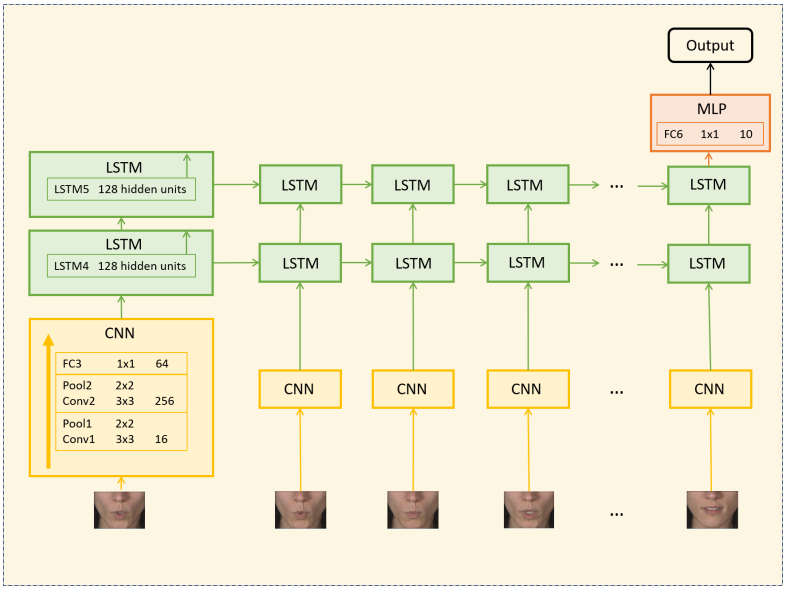

Figure S3: Architecture from Lee et al. [128]

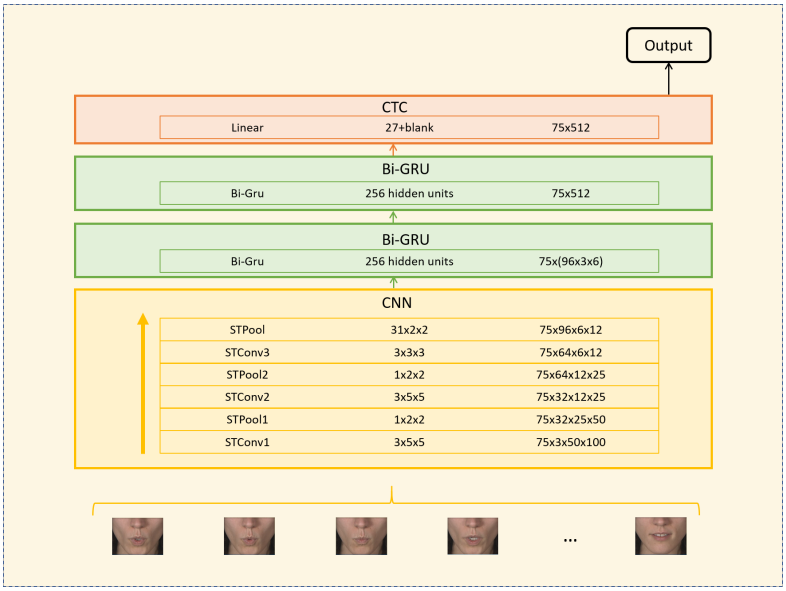

Figure S4: Architecture from Assael et al. [34] - LIPNET

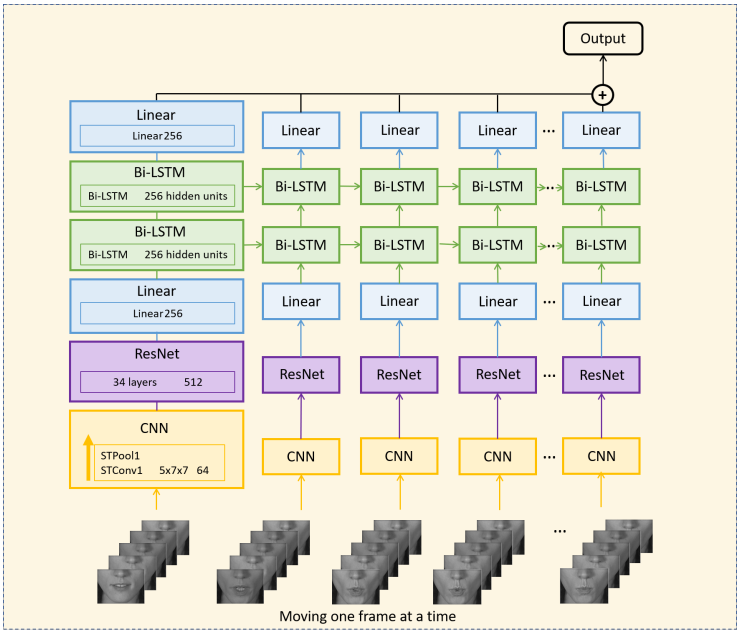

Figure S5: Architecture from Stafylakis et al. [156]

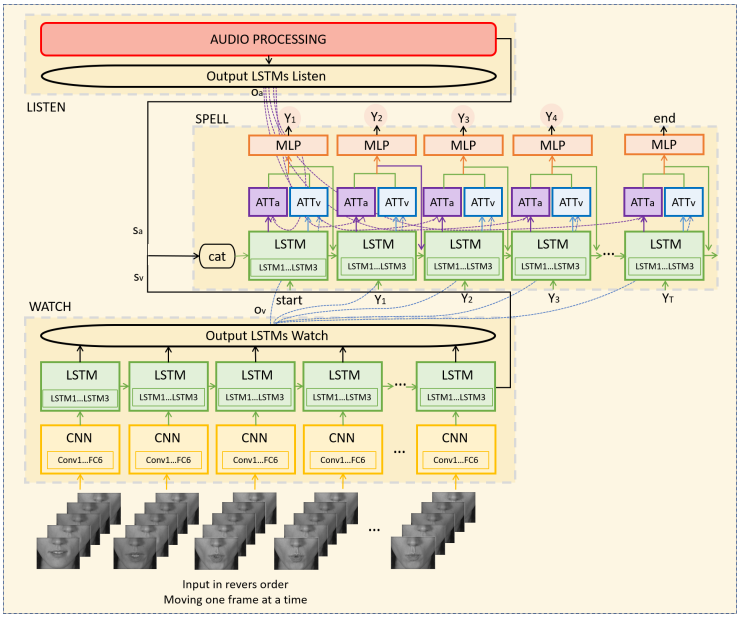

Figure S6: Architecture from Chung et al. [16] - WLAS

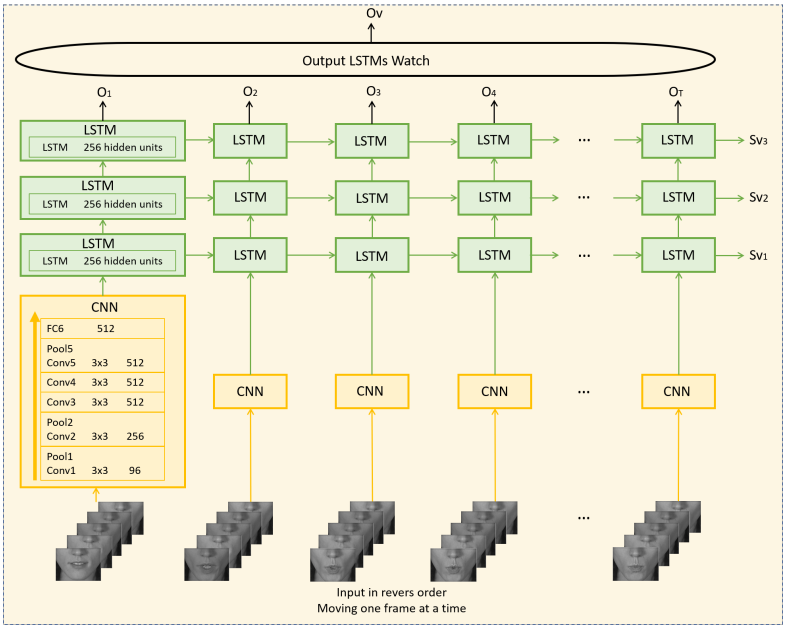

Figure S7: Architecture from Chung et al. [16] - WATCH

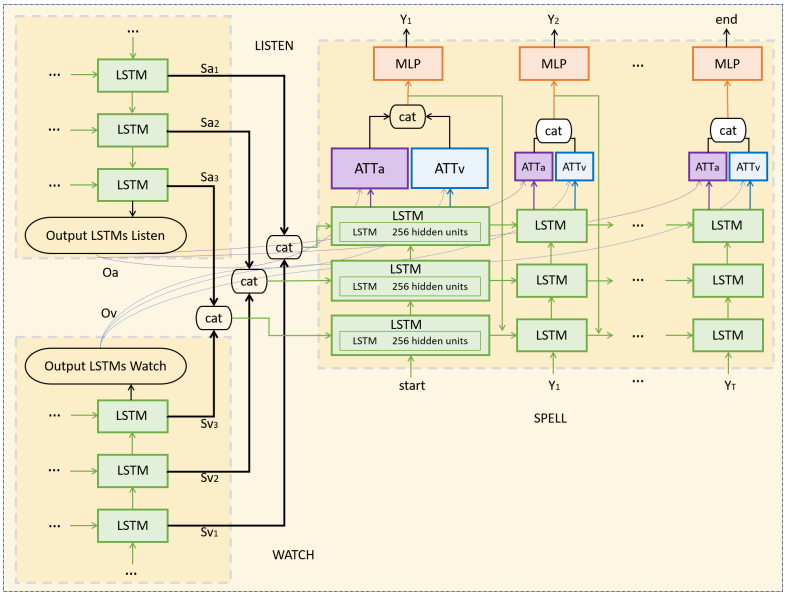

Figure S8: Architecture from Chung et al. [16] - SPELL

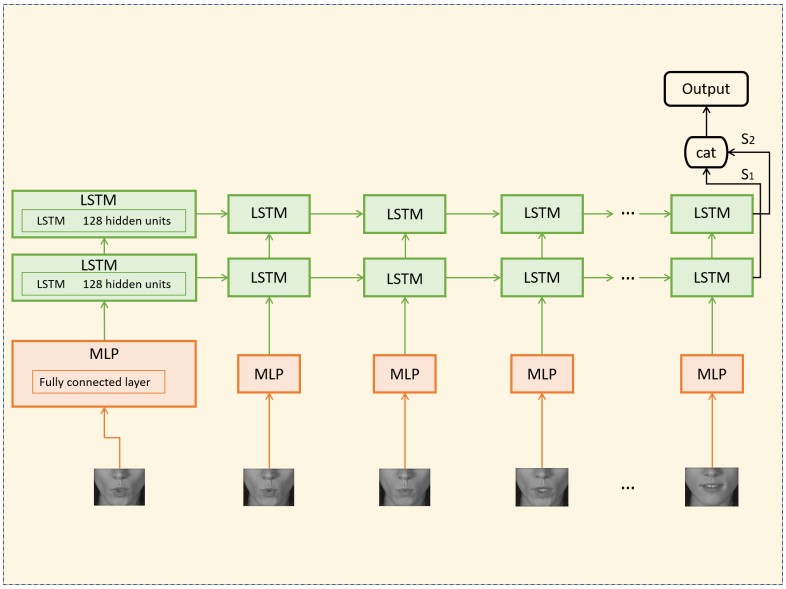

Figure S9: Architecture from Wand et al. [20]

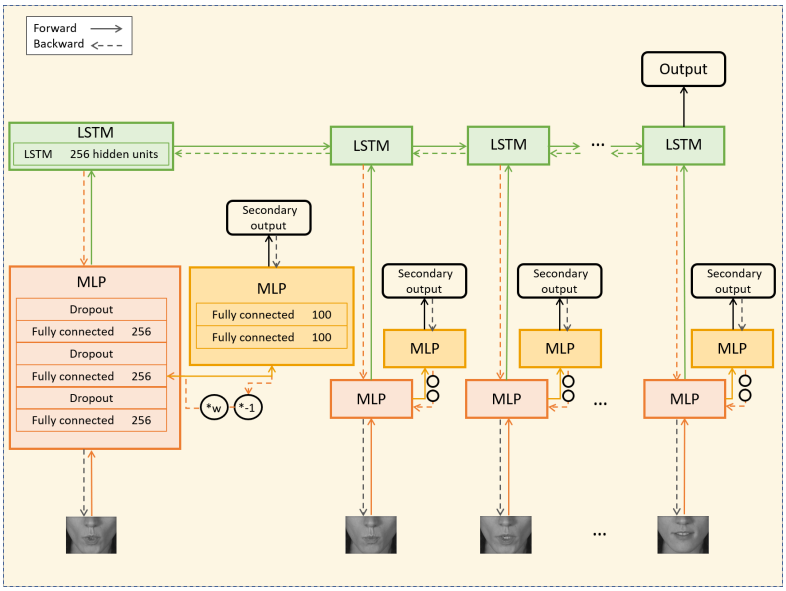

Figure S10: Architecture from Wand et al. [160]

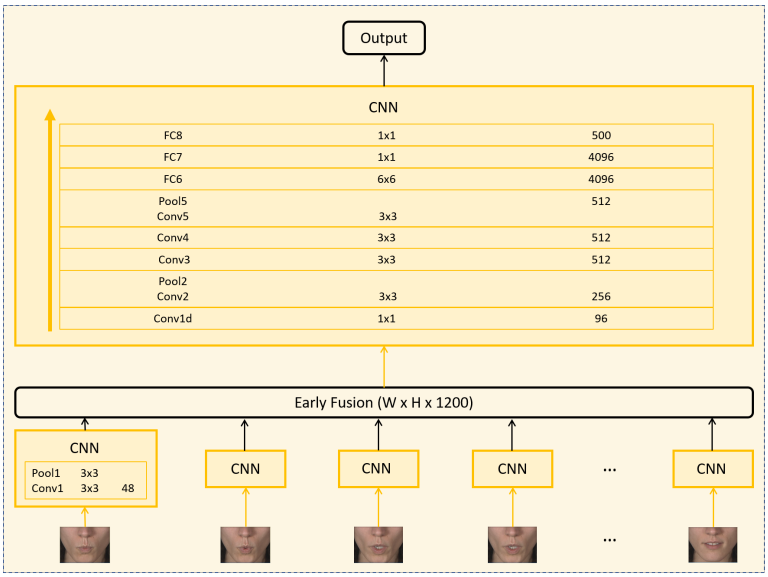

Figure S11: Architecture from Chung et al. [19]

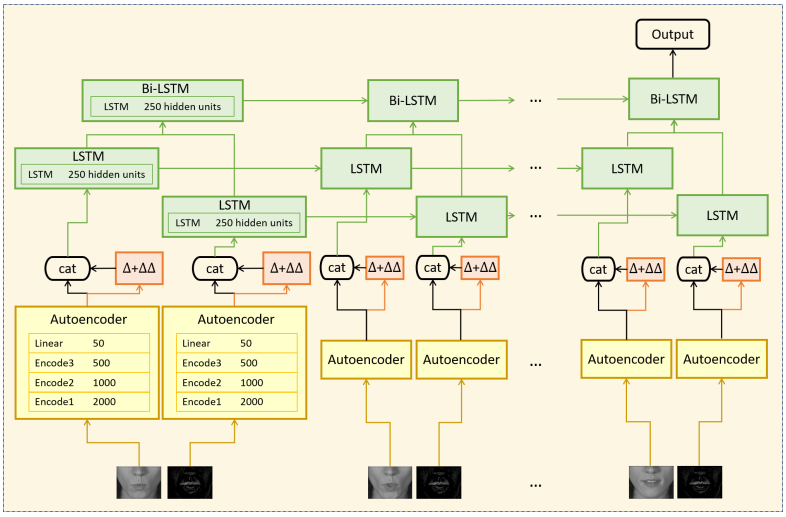

Figure S12: Architecture from Petridis et al. [154]