

ABSTRAK

Pembacaan gerak bibir adalah metode untuk memahami apa yang dikatakan pembicara dengan membaca gerakan bibir. Ini juga disebut pengenalan ucapan secara visual. Pengenalan ucapan visual akan menghasilkan teks dari video tanpa adanya audio. Teks tersebut terdiri dari beberapa kata atau bahkan kalimat yang diucapkan oleh pembicara. Tantangan yang sering dihadapi dalam pembacaan gerak bibir adalah variasi dalam input, yang membuat pengenalan ucapan menjadi lebih sulit. Variasi seperti fitur wajah dan kecepatan bicara yang berbeda dapat mengurangi akurasi. Saat ini, *deep learning* memberikan hasil yang menjanjikan dalam mengekstraksi fitur visual. Ini juga memberikan kesempatan bagi pembacaan gerak bibir untuk memberikan hasil yang lebih baik. Agar dapat menggunakan video sebagai input, penelitian ini menggunakan arsitektur 3D *Deep Learning*. Selain itu, keluar dari masalah kosakata juga membuat pengenalan ucapan visual lebih sulit untuk diterapkan di dunia nyata. Sistem hanya dapat memprediksi kata-kata yang ada di kamus. Itu sebabnya, penelitian ini menggunakan basis suku kata untuk menangani masalah tersebut. Dengan membuat sistem berbasis suku kata, ini memberikan kesempatan untuk membangun kata baru yang tidak ada dalam kamus. Kombinasi suku kata yang ada, digunakan untuk membangun kata baru. Karena data yang diperoleh terlalu kecil untuk *deep learning*, proses augmentasi dilakukan sebanyak 40 kali. Hasilnya, data setelah augmentasi dapat mencapai akurasi sebesar 100% dalam proses pengujian. Dalam 10 kata OOV yang dicoba, 8 di antaranya mampu dikenali oleh sistem.

Kata kunci: pembacaan bibir, pengenalan ucapan visua, deep learning, out of vocabulary