INTISARI

PENGEMBANGAN ARSITEKTUR TRANSFORMER DENGAN DYNAMIC ATTENTION MECHANISM UNTUK PENGENALAN BAHASA ISYARAT

Oleh

Elva Amalia

23/528841/PPA/06690

Bahasa isyarat memiliki variasi gerakan yang kompleks dan banyak di antaranya menunjukkan *interclass similarities*. Kondisi ini menyulitkan model pengenalan bahasa isyarat dalam membedakan gestur yang hampir serupa, terutama karena mekanisme *self-attention* konvensional cenderung memberikan bobot secara statis dan kurang responsif terhadap fitur spasial-temporal yang dinamis. Penelitian ini mengusulkan arsitektur Multimodal Transformer dengan mengganti *self-attention* menjadi *dynamic attention* berbasis *cosine similarity* untuk meningkatkan sensitivitas model terhadap perbedaan gerakan halus. Model memanfaatkan kombinasi fitur S3D dan keypoints yang diproses melalui lapisan Transformer sehingga informasi pose dan dinamika gerakan dapat direpresentasikan secara lebih diskriminatif. Evaluasi dilakukan pada dataset WLASL *interclass similarities* 100 dengan mengukur metrik akurasi, presisi, recall, dan F1-score baik pada keseluruhan kelas maupun subset kelas dengan *interclass similarities*. Hasil pengujian menunjukkan bahwa model dengan dynamic attention mampu meningkatkan akurasi global dari 0.55 menjadi 0.59. Selain itu, pada subset 20 kelas *interclass similarities*, performa juga meningkat dari 82.6% menjadi 84.0%.

Kata Kunci: Pengenalan Bahasa Isyarat, Transformer, Dynamic *Attention*, Interclass Similarities, WLASL

ABSTRACT

TRANSFORMER ARCHITECTURE DEVELOPMENT WITH DYNAMIC ATTENTION MECHANISM FOR SIGN LANGUAGE RECOGNITION

By

Elva Amalia

23/528841/PPA/06690

Sign language exhibits complex movement variations, many of which demonstrate high interclass similarity. This characteristic poses a challenge for sign language recognition models, as distinguishing between gestures with highly similar visual patterns becomes difficult, particularly because conventional self-attention mechanisms assign static weights and are less responsive to dynamic spatio-temporal features. This study proposes a Multimodal Transformer architecture that replaces the standard self-attention mechanism with cosine-similarity-based dynamic attention to enhance the model's sensitivity to subtle motion distinctions. The model employs a combination of S3D and keypoint features processed through Transformer layers, allowing both pose information and motion dynamics to be represented in a more discriminative manner. The evaluation was conducted on the WLASL Interclass Similarities 100 dataset using accuracy, precision, recall, and F1-score metrics, assessed across the full class set and a subset of classes with high interclass similarity. The experimental results demonstrate that the integration of dynamic attention improves global accuracy from 0.55 to 0.59. Additionally, performance within the subset of 20 interclass-similarity classes increased from 82.6% to 84.0%.

Keywords: Sign Language Recognition, Transformer, Dynamic *Attention*, Interclass Similarities, WLASL