

Abstrak

Teknik segmentasi dapat dimanfaatkan untuk membantu menyelesaikan masalah kehidupan sehari-hari. Pada bidang medis sendiri, teknik segmentasi semantik dapat diterapkan untuk membantu mendeteksi penyebaran kanker pada pasien kanker secara dini. Dalam bidang segmentasi semantik, pendekatan-pendekatan berbasis *Convolutional Neural Network* (CNN), seperti *Fully Convolutional Network* (FCN) dan *DeepLabv3+*, telah dikenal mendominasi pada bidang tersebut. Di samping itu, keberhasilan dari pendekatan *Transformer* pada bidang pemrosesan bahasa alami telah menarik banyak perhatian peneliti-peneliti untuk ikut terjun memanfaatkan pendekatan tersebut dalam menyelesaikan permasalahan segmentasi semantik. Hal itu memicu dikembangkannya *Vision Transformer* (ViT) sebagai alternatif baru dalam penyelesaian segmentasi semantik. Berbeda dengan pendekatan konvolusi yang menggunakan kernel yang digeser untuk mendapatkan informasi kontekstual lokal, pendekatan ViT menerima citra sebagai *patch* yang nantinya dapat digunakan untuk menghasilkan informasi kontekstual lokal serta global. Salah satu model yang terinspirasi dari arsitektur ViT yaitu *SegFormer* menggabungkan *Hierarchical Transformer Encoder* untuk menghasilkan fitur halus dengan resolusi rendah dan fitur kasar dengan resolusi tinggi dan dekoder *All-MLP* yang ringan untuk menggabungkan fitur *multi-level* yang dihasilkan dari enkoder untuk membuat *mask* segmentasi terakhir. Pada penelitian ini, model *SegFormer* digunakan untuk melakukan segmentasi terhadap citra pemindaian tulang seluruh tubuh. Hasilnya, dengan membandingkan model *SegFormer* dengan beberapa model konvolusi yaitu FCN dan *DeepLabv3+*, performa dari *SegFormer* berhasil mengalahkan kedua model konvolusi tersebut dengan nilai mIoU tertinggi yang diraih yaitu 77,86% pada data tes.

Kata kunci: citra pindai tulang, segmentasi semantik, *Transformer*, konvolusi, *SegFormer*, *DeepLabv3+*, *FCN*