BAB I
PENDAHULUAN

1.1 Latar Belakang

Actions atau aksi merupakan aktivitas seseorang yang terdiri dari beberapa gerakan yang teratur dan memiliki makna seperti berjalan, melambai, dan meninju (Aggarwal & Ryoo, 2007). Manusia dapat dengan mudah mengenali dan mengidentifikasi suatu tindakan atau aksi seseorang dalam sebuah video, namun tidak demikian dengan mesin. Memahami perilaku manusia dalam video tetap menjadi masalah utama dalam *computer vision* (De Souza, 2018). Hal ini sulit dilakukan karena beberapa hal, seperti adanya variasi dalam melakukan suatu aksi, objeknya dapat berubah, atau sudut pandang yang berbeda (Subroto & Liliana, 2021). Teknik *action recognition* dapat menganalisis isi video untuk dapat menentukan tindakan apa saja yang ada pada video tersebut. *Action recognition* ini dapat dimanfaatkan di berbagai bidang seperti *security surveillance, autonomous driving, robotics* (De Souza, 2018), *early warning* (Subroto & Liliana, 2021), *video retrieval* (Sch & Barbara, 2004), dan *healthcare* (Murad & Pyun, 2017; Ogbuabor & La, 2018).

Terdapat beberapa penelitian mengenai *action recognition* yang menggunakan berbagai metode berbeda. Penelitian sebelumnya menggunakan metode *Support Vector Machine* (SVM) dan *Bag of Word* (BOW) menghasilkan akurasi 70,37% dengan dataset UTD-MHAD versi video RGB (Ben Mahjoub & Atri, 2016). Kemudian penelitian dengan dataset yang sama (UTD-MHAD) menggunakan metode *Dynamic Texture Image* (DTI) menghasilkan akurasi 79,81% (Liu et al., 2019). Eun dkk (2020) menggunakan metode *Recurrent Neural Networks* (RNN) dengan model *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) pada penelitian *online action recognition* untuk serial TV, dan berhasil mendapatkan akurasi 80,9% untuk RNN-LSTM dan 81,3% untuk RNN-GRU. Penelitian lain menggunakan *Deep Recurrent Neural Networks* (DRNNs) dengan LSTM mampu menangkap dependesi jarak jauh dalam urutan input panjang variabel, berhasil mencapai akurasi 96.7% untuk dataset UCI (Murad

& Pyun, 2017). Penggunaan *Convolutional Neural Network Classifier* (CNN) dengan representasi fitur bilangan fuzzy mencapai akurasi 97.8% (Mohmed et al., 2020).

Disisi lain, pada bidang *Natural Language Processing* (NLP) terdapat pendekatan *State-of-the-Art* yaitu metode Tranformer (Wolf et al., 2020), metode ini mampu menangani *Sequence to Sequence* pada bidang NLP. Transformer telah memberikan hasil yang memuaskan dalam berbagai masalah pada bidang NLP. Salah satu inovasi terkini dalam Transformer adalah pemanfaatan *Generated pre-trained Transformer* generasi ke-3 (GPT3) yang mendapat penghargaan Paper Terbaik di Konferensi *Neural Information Processing Systems* tahun 2020 atau lebih dikenal dengan NeurIPS2020 (Chen et al., 2021). Transformer juga tidak mengharuskan data sekuensial diproses secara berurutan, sehingga metode ini dapat memberikan waktu pelatihan yang lebih sedikit dibandingkan RNN (Zhai et al., 2021).

Sebetulnya *action recognition* ini dapat dianalogikan sebagai kumpulan *frame* seperti halnya pada NLP yang merupakan *sequence* kata (Cho & Foroosh, 2019). Oleh karena itu pada penelitian ini diimplementasikan Transformer untuk menyelesaikan masalah *action recognition*. Walaupun sejak tahun 2012 dalam bidang *computer vision*, *Convolutional Neural Network* (CNN) telah menjadi model yang dominan untuk tugas-tugas *computer vision* (Zhai et al., 2021), namun menggunakan Transformers untuk *computer vision* menjadi arah penelitian baru, karena dapat mengurangi kompleksitas arsitektur (Chen et al., 2021). Maka dari itu, diharapkan penggunaan metode ini mampu menghasilkan hasil unjuk kerja yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan tersebut maka dirumuskan masalah sebagai berikut:

1) Bagaimana implementasi metode Transformer untuk pengenalan aksi pada video?
2) Bagaimana unjuk kerja metode Transformer untuk pengenalan aksi pada video?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang terlah diuraikan di atas, tujuan penelitian ini, yaitu:

1) Mengimplementasi metode Transformer untuk pengenalan aksi pada video.
2) Menganalisis unjuk kerja metode Transformer untuk pengenalan aksi pada video.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah aksi-aksi yang terdapat pada data yang digunakan, yaitu set data publik *Multimodal Human Action Detection* dari University of Texas at Dallas atau selanjutnya disebut UTD-MHAD dengan jenis video RGB untuk pengenalan aksi manusia.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini sebagai berikut:

1) Bagi Peneliti

Peneliti diharapkan mampu mendapatkan pengetahuan baru mengenai penyelesaian masalah *action recognition* pada video. Selain itu juga peneliti mampu memahami proses dan alur kerja dari penggabungan metode pada model *deep learning*.

2) Bagi Pihak Lain

Hasil penelitian ini diharapkan mampu diimplementasikan dalam penyelesaian masalah *action recognition* pada video dan dapat dijadikan rujukan pada penelitian selanjutnya.

1.6 Sistematika Penulisan

Adapun sistematika penulisan karya ilmiah yang berisi mengenai gambaran umum dan format skripsi, adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang latar belakang penelitian mengapa *action recognition* pada video ini diusulkan. Kemudian, bab ini menguraikan rumusan masalah dan tujuan penelitian mengenai implementasi dan unjuk kerja dari implementasi metode Transformer terhadap *action recognition*. Selain itu, terdapat batasan masalah dan manfaat dalam penelitian *action recognition* ini. Serta sistematika penulisan yang menjelaskan mengenasi isi dari semua bab yang terdapat pada penelitian.

BAB II KAJIAN PUSTAKA

Sebelum melakukan penelitian tentang *action recognition*, diperlukan pemahaman teori dan metode yang dibutuhkan untuk merancang dan membangun model *action recognition* menggunakan metode Transformer. Adapun teori dan metode yang dipaparkan pada bab ini terdiri dari *computer vision*, pengolahan citra digital, *action recognition*, dan metode Transformer. Selain itu terdapat konsep dan

gagasan terhadap studi kasus yang diangkat dari berbagai penelitian sebelumnya yang relevan dengan *action recognition*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tahapan-tahapan yang dilakukan pada penelitian *action recognition*, antara lain desain penelitian yang menjabarkan bagaimana alur penelitian dari mulai perumusan masalah mengapa penelitian *action recognition* dengan metode Transformer ini dilakukan, kemudian dilanjutkan dengan studi literatur untuk mendukung teori dan metode dari *action recognition*, pengumpulan data, dan desain eksperimen dari mulai *input* video, *praprocessing*, *training*, *testing*, dan evaluasi hasil. Hingga tahap terakhir yaitu penarikan kesimpulan. Pada tahapan tersebut pun dijelaskan mengenai rencana implementasi yang berisi tentang usulan beberapa skenario *training* yang selanjutnya akan diimplementasikan dalam eksperimen *action recognition*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang uraian hasil dan pembahasan mengenai hasil implementasi penelitian *action recognition* yang telah dilakukan sesuai dengan desain eksperimen yang ada pada bab 3, mulai dari hasil *praprocessing input* data, *construct model*, hasil *training* dan *testing*, sampai dengan evaluasi dan analisis model. Bab ini juga menyertakan analisis dari hasil eksperimen yang dilakukan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari penelitian *action recognition* yang telah dilakukan sesuai dengan rumusan masalah mengapa penelitian ini dilakukan, dan bagaimana implementasi metode Transformer dalam bidang *action recognition*, disertai dengan saran rekomendasi mengenai pengembangan untuk penelitian terkait berikutnya.