

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi yang pesat telah menyebabkan peningkatan eksponensial dalam volume gambar yang dihasilkan dan dibagikan di internet. Menurut data yang diambil dari situs Mylio, sekitar 1,5 triliun foto diambil pada tahun 2022 dan diperkirakan akan mencapai 1,98 triliun pada tahun 2026 [1]. Peningkatan volume gambar yang signifikan ini menciptakan kebutuhan terhadap sistem pencarian dan pengambilan gambar yang efisien serta akurat [2].

Metode pengambilan gambar tradisional pada umumnya dapat diklasifikasikan ke dalam dua kategori yaitu *Text-Based Image Retrieval* (TBIR) dan *Content-Based Image Retrieval* (CBIR). TBIR mengandalkan kata kunci dan deskripsi yang dibuat oleh pengguna untuk mengindeks dan mencari gambar, sedangkan CBIR menggunakan fitur visual seperti warna, tekstur, dan bentuk untuk menemukan gambar yang serupa. Namun, kedua metode tersebut memiliki keterbatasan. TBIR sering kali memiliki kata kunci yang tidak lengkap dan ambigu[3], sementara CBIR kesulitan untuk menghubungkan celah semantik antara fitur visual tingkat rendah dan konsep tingkat tinggi[4].

Teknik *Semantic-based image retrieval* (SBIR) bertujuan untuk mengatasi keterbatasan TBIR dan CBIR dengan mengindeks dan mencari gambar berdasarkan makna semantiknya. Teknik SBIR menggunakan pemrosesan bahasa alami dan analisis semantik untuk memahami makna dari gambar, sehingga memungkinkan untuk melakukan pencarian gambar berdasarkan konsep atau objek yang terkandung di dalamnya[5]. Teknik SBIR dapat membantu mengatasi masalah ketidakakuratan dan ketidakjelasan hasil pencarian pada TBIR dan CBIR. Sebagai contoh, pengguna yang ingin mencari "burung yang sedang duduk di dahan pohon" akan diperlihatkan gambar-gambar burung yang sedang duduk di dahan pohon, bukan hanya gambar-gambar yang kebetulan ditandai dengan kata-kata tersebut.

Salah satu solusi yang bisa digunakan untuk SBIR adalah dengan menggunakan model OpenAI *Contrastive Language-Image Pre-training* (CLIP). CLIP adalah model deep learning yang telah dilatih sebelumnya pada sekumpulan besar dataset gambar dan deskripsi tekstual yang terkait[6]. CLIP dapat menghasilkan fitur visual dan tekstual yang bermakna yang dapat digabungkan

untuk meningkatkan kinerja pengambilan gambar. Dengan menggunakan CLIP untuk SBIR, aplikasi pencarian gambar dapat secara efektif mengambil gambar yang relevan berdasarkan konten visual dan konteks semantiknya.

Pada penelitian yang dilakukan sebelumnya oleh [7], model CLIP diterapkan untuk mencari video yang mengandung *bug* dalam konteks *video game*. Penelitian tersebut menggunakan dua varian model CLIP, yaitu "RN101" dan "ViT-B/32". Dataset yang digunakan terdiri dari video yang berasal dari subreddit GamePhysics. Untuk mengukur kesamaan antara representasi video dan teks yang dihasilkan oleh CLIP, digunakan metrik *cosine similarity*. Hasil penelitian ini menunjukkan bahwa dari 6.192 video dari berbagai *game*, akurasi rata-rata yang dicapai adalah 66,24% untuk *query bug* dengan metrik recall@5 pada berbagai arsitektur CLIP.

Adapun dalam penelitian ini, berfokus pada pengembangan aplikasi yang memanfaatkan kemampuan CLIP untuk tujuan pencarian gambar. *Cosine similarity* digunakan sebagai teknik pengukuran untuk menentukan kesamaan antara dua vektor. Dalam konteks mesin pencari gambar, salah satu pendekatan adalah mengubah baik gambar maupun teks permintaan menjadi vektor menggunakan CLIP, dan kemudian menghitung *cosine similarity* di antara keduanya. Oleh karena itu, tujuan dari penelitian ini adalah mengimplementasikan CLIP dalam aplikasi pencarian gambar, memanfaatkan *cosine similarity*, dan mengevaluasi efektivitasnya menggunakan dataset Microsoft Common Objects in Context (MSCOCO) [8] untuk menilai kemampuan CLIP dalam menghubungkan teks dan gambar.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, berikut adalah rumusan masalah penelitian yang diajukan :

1. Bagaimana cara implementasi OpenAI CLIP dalam aplikasi pencarian gambar?
2. Berapa tingkat akurasi dari OpenAI CLIP menggunakan *cosine similarity*?

1.3 Batasan Permasalahan

Berikut adalah batasan masalah :

1. Penelitian ini akan menggunakan model pretrained clip-ViT-B-32 tanpa melakukan training lebih lanjut.
2. Penelitian ini akan menggunakan bahasa Inggris sebagai bahasa yang digunakan dalam teks *query*.
3. Dataset yang akan digunakan untuk pengujian adalah dataset MSCOCO 2017.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan OpenAI CLIP pada aplikasi pencarian gambar.
2. Mengukur tingkat akurasi OpenAI CLIP menggunakan *cosine similarity*.

1.5 Manfaat Penelitian

Manfaat dari Penelitian ini adalah:

1. Menjadi referensi bagi pengembang aplikasi pencarian gambar dalam mengimplementasikan CLIP pada aplikasi mereka.
2. Menambah ilmu pengetahuan.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- Bab 2 LANDASAN TEORI

Bab ini menjelaskan teori-teori yang digunakan pada penelitian seperti OpenAI CLIP, dan *Cosine Similarity*.

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan pada penelitian. Adapun metode yang dimaksud adalah studi literatur, perancangan aplikasi, pembuatan aplikasi, pengujian dan evaluasi, dan penulisan laporan.

- Bab 4 HASIL DAN DISKUSI

Bab ini menjelaskan hasil implementasi yang dilakukan berdasarkan metode penelitian serta hasil uji coba dan evaluasi.

- Bab 5 SIMPULAN DAN SARAN

Bab ini menjelaskan kesimpulan dari penelitian yang sudah dilakukan dan saran untuk mengembangkan penelitian ini.

