

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam dunia menulis, inspirasi seringkali datang dari berbagai sumber, termasuk elemen visual dan gambar yang mampu memicu imajinasi penulis untuk menciptakan karya yang menarik [1], [2], [3]. Perkembangan teknologi, khususnya dalam bidang kecerdasan buatan, telah membawa perubahan signifikan terhadap gaya menulis dan pola komunikasi para penulis [4]. Berbagai sistem telah dikembangkan untuk mengolah konten visual menjadi deskripsi atau narasi, memungkinkan penulis mendapatkan bantuan dalam membangun ide. Meskipun demikian, fenomena *writer's block* yang merupakan sebuah kondisi yang menyebabkan penulis mengalami kebuntuan ide masih menjadi masalah umum, bahkan di era ketika alat bantu berbasis kecerdasan buatan semakin meluas [5], [6], [7], [8]. Kebuntuan ini dapat menghambat penulis dalam mengasah kepekaan visual dan kemampuan naratif, membuat mereka kesulitan untuk menangkap esensi dari apa yang mereka lihat dan mengubahnya menjadi kata-kata yang memikat [9].

Salah satu penelitian yang mendalami fenomena *writer's block* adalah tesis oleh S. J. Ahmed berjudul "An analysis of writer's block: causes, characteristics, and solutions" dari University of North Florida [10]. Melalui survei yang dilakukan terhadap 146 penulis dengan pengalaman di berbagai genre fiksi dan nonfiksi, studi tersebut menemukan bahwa mayoritas penulis secara konsisten mengalami kebuntuan ide, dengan tingkat reliabilitas statistik yang tinggi (Cohen's Kappa 0,77–0,90). Temuan ini mengkonfirmasi bahwa *writer's block* bukanlah keluhan biasa, melainkan sebuah masalah valid dan terukur dalam komunitas kreatif profesional, yang seringkali dipicu oleh ketidakpastian dalam memulai atau mengembangkan arah cerita. Hal ini menegaskan bahwa *writer's block* adalah masalah umum yang mempengaruhi banyak penulis dan dapat berdampak signifikan pada produktivitas serta kreativitas mereka.

Menyadari kompleksitas *writer's block* sebagai fenomena psikologis, penelitian ini tidak mengklaim untuk dapat menyelesaikannya secara menyeluruh. Sebaliknya, penelitian ini berfokus pada salah satu aspeknya yang paling umum dan

dapat diatasi secara teknis, yaitu kesulitan untuk memulai proses menulis dari halaman kosong (*blank page syndrome*). Untuk menjembatani masalah ini dengan solusi yang konkret, aplikasi yang diusulkan diposisikan sebagai alat bantu kreativitas praktis yang unik, karena output yang dihasilkan bukanlah deskripsi harfiah, melainkan sebuah narasi yang dirangkai secara puitis dan kaya akan majas. Sebagai contoh, seorang kreator konten edukasi ingin membuat materi tentang alam. Alih-alih mendapatkan deskripsi faktual seperti "ini adalah gambar gunung dengan pepohonan", aplikasi ini bertujuan menghasilkan titik awal yang lebih imajinatif. Misalnya, dari gambar yang sama, output yang diharapkan adalah narasi seperti: "Punggung sang raksasa batu tertidur lelap di bawah selimut kabut pagi, sementara pepohonan berbaris laksana prajurit setia yang menjaga tidurnya." Draf puitis inilah yang menjadi keunikan solusi yang ditawarkan, berfungsi sebagai pemicu kreativitas yang dapat dikembangkan lebih lanjut oleh pengguna, baik untuk tujuan edukasi, artistik, maupun komersial.

Oleh karena itu, dibutuhkan pendekatan baru yang tidak hanya menerjemahkan gambar menjadi teks, tetapi juga mampu menghadirkan inspirasi naratif secara lebih relevan terhadap konteks visual yang dianalisis. Penggunaan gambar sebagai sumber inspirasi dapat menjadi cara yang efektif untuk merangsang kreativitas [11]. Gambar sering kali mengandung elemen visual yang kaya, yang dapat membangkitkan berbagai emosi dan imajinasi penulis. Penelitian ini mengusulkan sebuah *pipeline* yang mengintegrasikan modul deteksi objek visual menggunakan *Convolutional Neural Networks* (CNN) dengan modul generasi teks naratif berbasis *Generative Pre-trained Transformer 2* (GPT-2), yang secara khusus diarahkan untuk menghasilkan gaya bahasa sastra [12], [13]. Pemilihan arsitektur ini didasarkan pada pertimbangan metodologis yang strategis. Meskipun terdapat arsitektur deteksi objek yang lebih modern seperti YOLO (*You Only Look Once*), yang unggul dalam kecepatan dan akurasi, fokus utama dan kebaruan (*novelty*) penelitian ini tidak terletak pada optimasi performa deteksi objek itu sendiri [14], [15]. Sebaliknya, kontribusi utama penelitian ini terletak pada perancangan dan eksplorasi sebuah "jembatan" konseptual: yaitu, bagaimana output terstruktur dari sebuah detektor objek (berupa daftar entitas visual) dapat secara efektif diubah menjadi prompt yang kaya konteks untuk memicu imajinasi sebuah

model bahasa [16], [17]. Dengan demikian, CNN dipilih sebagai arsitektur dasar yang representatif untuk menyediakan input visual, memungkinkan penelitian untuk lebih fokus pada tantangan utama, yaitu proses transformasi dari 'daftar objek' menjadi 'alur cerita'.

Di sisi generasi teks, GPT-2 dipilih karena kemampuannya yang terbukti dalam menghasilkan teks yang tidak hanya relevan secara semantik tetapi juga koheren dan kreatif. Secara spesifik, arsitektur Transformer dan data pra-pelatihan skala besar yang mencakup karya sastra memungkinkan GPT-2 untuk mempelajari dan meniru pola linguistik yang kompleks, termasuk penggunaan gaya bahasa dan majas [18], [19]. Kemampuan inilah yang dieksploitasi dalam penelitian ini untuk menghasilkan teks yang melampaui deskripsi harfiah. Sinergi antara kedua model ini menjadi inti dari penelitian: CNN bertugas menyediakan "kerangka" logis dari gambar (objek-objek yang ada), sementara GPT-2 mengambil kerangka tersebut dan "melukisnya" dengan kata-kata puitis, memberikan 'nyawa' dan 'rasa' pada pemandangan visual.

Dukungan empiris untuk efektivitas penggunaan CNN dan *transformer* GPT-2 dalam pembuatan ide cerita dari gambar visual dapat ditemukan dalam penelitian terbaru. Sebagai contoh, artikel penelitian yang berjudul "Image Captioning with Audio Reinforcement Using RNN and CNN" menunjukkan bahwa kombinasi antara ekstraksi fitur gambar yang akurat menggunakan CNN dan kemampuan generasi teks yang canggih dari jaringan GPT-2 dapat menghasilkan deskripsi tulisan menjadi *visual storytelling*, di mana sistem tidak hanya mendeskripsikan, tetapi juga menarasikan sebuah cerita berdasarkan input visual relevan dari gambar-gambar [20]. Penelitian ini bertujuan untuk berkontribusi pada domain tersebut dengan tujuan praktis yang jelas: menyediakan alat yang dapat berfungsi sebagai katalis ide, khususnya bagi para pendidik, pembuat konten, atau bahkan penulis pemula yang membutuhkan percikan inspirasi untuk memulai proyek naratif mereka. Dengan mengintegrasikan teknologi CNN dan GPT-2, aplikasi ini memungkinkan pengguna mengubah gambar statis menjadi landasan cerita yang imajinatif dan kaya secara gaya bahasa. Dengan demikian, kehadiran aplikasi ini tidak hanya memperkaya pengalaman penulisan, tetapi juga menawarkan solusi

aplikatif untuk mengatasi tantangan awal dalam pengembangan narasi yang tidak hanya fungsional tetapi juga bernilai estetis.

Berdasarkan pemaparan di atas, judul yang diusulkan adalah "Transformasi Ide Cerita dari Gambar ke Teks Menggunakan *Convolutional Neural Networks* (CNN) dan *Generative Pre-trained Transformer* (GPT-2)".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, terdapat beberapa rumusan masalah yang dapat diidentifikasi terkait dengan proses transformasi ide cerita dari gambar ke dalam teks, yaitu:

1. Bagaimana cara mengimplementasikan teknik CNN dan transformer GPT-2 untuk membuat program yang dapat menghasilkan teks naratif dari gambar?
2. Bagaimana tingkat akurasi hasil teks naratif yang dihasilkan dari penggabungan teknik CNN dan GPT-2 dalam menginterpretasikan isi gambar?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian agar tidak terlalu meluas dari pembahasan yang dimaksud, penelitian ini membatasi ruang lingkupnya. Oleh karena itu, batasan masalah dalam penelitian ini meliputi:

1. Penelitian ini menggunakan gambar digital pemandangan alam sebagai input dan difokuskan hanya pada 7 kategori objek alam yang telah ditentukan, yaitu gunung, sungai, pohon, danau, langit, laut dan batu. Dengan demikian, sistem tidak dirancang untuk memproses atau mengenali objek dari jenis gambar lain maupun objek alam di luar 7 kategori yang telah ditetapkan.
2. Output yang dihasilkan sistem adalah narasi pendek (1-3 kalimat) dalam Bahasa Indonesia yang bersifat puitis dan interpretatif, bukan deskripsi faktual atau cerita panjang.

1.4 Tujuan Penelitian

Dengan mempertimbangkan kompleksitas proses transformasi ide cerita dari visual ke teks, tujuan penelitian ini adalah untuk menghasilkan solusi inovatif yang

dapat mengatasi tantangan tersebut. Berikut adalah tujuan-tujuan yang hendak dicapai dalam penelitian ini:

1. Mengembangkan program yang mengimplementasikan teknik CNN dan *transformer* GPT-2 untuk menghasilkan teks naratif dari gambar.
2. Menguji tingkat akurasi teks naratif yang dihasilkan oleh penggabungan teknik CNN dan GPT-2 dalam merepresentasikan isi gambar.

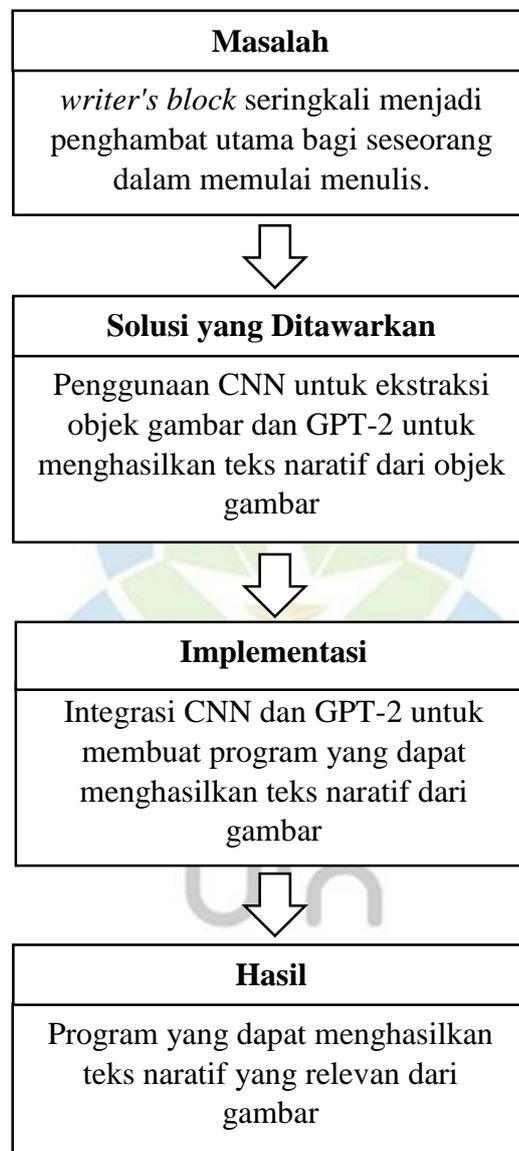
1.5 Kerangka Pemikiran

Kerangka pemikiran penelitian ini diawali dengan identifikasi masalah fundamental dalam proses kreatif, yaitu kesenjangan antara persepsi visual yang kaya akan potensi cerita dengan kesulitan dalam menuangkannya menjadi teks, sebuah tantangan yang sering disebut sebagai *writer's block*. Untuk menjembatani kesenjangan ini, penelitian ini mengusulkan sebuah kerangka kerja berbasis kecerdasan buatan. Solusi yang ditawarkan adalah dengan mengintegrasikan dua model deep learning yang memiliki fungsi komplementer: Convolutional Neural Network (CNN) untuk analisis visual dan Generative Pre-trained Transformer 2 (GPT-2) untuk generasi narasi.

Secara teknis, alur kerja sistem dimulai saat sebuah gambar digital menjadi input. Modul pertama, yaitu CNN, berperan sebagai "mata" sistem yang menganalisis gambar untuk mendeteksi dan mengekstraksi fitur-fitur visual kunci, seperti objek-objek yang ada di dalamnya. Informasi terstruktur hasil deteksi objek ini kemudian diteruskan sebagai prompt atau pemicu kepada modul kedua, yaitu GPT-2. Berperan sebagai "otak" kreatif, GPT-2 akan mengolah prompt visual tersebut dan menghasilkan sebuah narasi pendek yang koheren, relevan, dan bergaya puitis sesuai dengan konteks gambar.

Tahap akhir dari kerangka pemikiran ini adalah implementasi dan validasi. Sistem yang telah dikembangkan akan diimplementasikan untuk diuji kinerjanya secara sistematis. Proses evaluasi akan dilakukan untuk mengukur efektivitas setiap modul—akurasi untuk CNN dan kualitas narasi untuk GPT-2—serta sistem secara keseluruhan. Hasil dari evaluasi ini diharapkan dapat memberikan wawasan mendalam tentang sejauh mana arsitektur yang diusulkan mampu menjadi solusi praktis dalam menyediakan stimulus kreatif dan membantu penulis mengatasi

hambatan awal dalam proses penulisan mereka. Berikut adalah bagan kerangka pemikiran untuk penelitian:



Gambar 1.1 Kerangka pemikiran penelitian

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang proses penulisan yang dilakukan untuk menyelesaikan penelitian ini. Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, perumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka pemikiran penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini terdiri dari pembahasan konsep dasar dan teori-teori yang berkaitan dengan topik masalah yang diambil dan hal-hal yang berguna dalam proses analisis permasalahan.

BAB III METODOLOGI PENELITIAN

Bab ini terdiri dari pembahasan tentang alur metodologi penelitian dari tahap awal sampai akhir yang di dalamnya terdiri dari pembahasan analisa produk dan perencanaan eksekusi aplikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini terdiri dari dua hal utama, pertama pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Pemaparan hasil penelitian disesuaikan dengan urutan masalah penelitian. Kedua pembahasan hasil atau temuan penelitian untuk menjawab rumusan penelitian.

BAB V PENUTUP

Bab ini terdiri dari pembahasan kesimpulan penelitian, kritik, dan saran yang penulis dapatkan selama menyelesaikan penelitian tugas akhir ini. Intisari dari bab ini dapat digunakan sebagai pertimbangan untuk pengembangan penulisan selanjutnya.