KERJA PRAKTIK – EF234603

PELABELAN DAN PEMBUATAN MODEL IMAGE CAPTIONING MENGGUNAKAN DEEP LEARNING

Laboratorium Komputasi Cerdas dan Visi (KCV)
Jalan Teknik Kimia - Gedung Departemen Teknik Informatika
Kampus Institut Teknologi Sepuluh Nopember Surabaya
Jalan Raya ITS, Sukolilo, Surabaya, Jawa Timur.

Oleh:
Wardatul Amalia Safitri 5025211006
Hammuda Arsyad 5025211146

Pembimbing Jurusan
Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc.

Pembimbing Lapangan
Dini Adni Navastara, S.Kom., M.Sc.

DEPARTEMEN TEKNIK INFORMATIKA
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2024

KERJA PRAKTIK - EF234603

PELABELAN DAN PEMBUATAN MODEL IMAGE CAPTIONING MENGGUNAKAN DEEP LEARNING

Laboratorium Komputasi Cerdas dan Visi (KCV)
Jalan Teknik Kimia - Gedung Departemen Teknik Informatika
Kampus Institut Teknologi Sepuluh Nopember Surabaya
Jalan Raya ITS, Sukolilo, Surabaya, Jawa Timur.

Oleh:

Wardatul Amalia Safitri	5025211006
Hammuda Arsyad	5025211146

Pembimbing Jurusan
Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc.

Pembimbing Lapangan
Dini Adni Navastara, S.Kom., M.Sc.

DEPARTEMEN TEKNIK INFORMATIKA
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2024

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

[Halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

[Halaman ini sengaja dikosongkan]

DAFTAR TABEL

[Halaman ini sengaja dikosongkan]

LEMBAR PENGESAHAN
KERJA PRAKTIK

PELABELAN DAN PEMBUATAN MODEL IMAGE CAPTIONING MENGGUNAKAN DEEP LEARNING

Oleh:

Wardatul Amalia Safitri	5025211006
Hammuda Arsyad	5025211146

Disetujui oleh Pembimbing Kerja Praktik:

1. Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc.
   NIP. 198106202005011003

   (Pembimbing Departemen)

2. Dini Adni Navastara, S.Kom., M.Sc.
   NIP. 198510172015042001

   (Pembimbing Lapangan)

[Halaman ini sengaja dikosongkan]

PELABELAN DAN PEMBUATAN MODEL IMAGE CAPTIONING MENGGUNAKAN DEEP LEARNING

Nama Mahasiswa	: Wardatul Amalia Safitri
NRP	: 5025211006
Nama Mahasiswa	: Hammuda Arsyad
NRP	: 5025211146
Departemen	: Teknik Informatika FTEIC-ITS
Pembimbing Departemen	: Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc.
Pembimbing Lapangan	: Dini Adni Navastara, S.Kom, M.Sc.

ABSTRAK

Kerja praktik ini dilakukan di Laboratorium Komputasi Cerdas dan Visi (KCV) Departemen Teknik Informatika Institut Teknologi Sepuluh Nopember (ITS). Kegiatan yang dilakukan memiliki tujuan untuk menghasilkan model image captioning dengan memanfaatkan deep learning. Kegiatan kerja praktik dimulai dengan membuat dataset citra trotoar dan lingkungan ITS beserta pelabelan atau pemberian deskripsi di setiap citra. Selanjutnya, dataset citra yang sudah diberi label akan diolah menjadi model image captioning dengan beberapa skenario implementasi deep learning. Metode deep learning yang diimplementasikan dalam kerja praktik ini meliputi LSTM, CNN, dan GRU. Hasil pengembangan model akan dibandingkan performanya menggunakan parameter BLEU Score dan ROUGE. Hasil evaluasi menunjukkan bahwa Dataset 1 menghasilkan performa terbaik dengan BLEU-1 51.49% dan ROUGE-1 40.05%. Metode CNN memberikan hasil terbaik dengan BLEU-1 25.56% dan ROUGE-1 23.44%. Sementara itu, fungsi aktivasi Relu

memberikan performa terbaik pada hyperparameter tuning dengan BLEU-1 23.61% dan ROUGE-1 22.37%.

Kata Kunci : Image Captioning, Deep Learning, LSTM, CNN, GRU

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia-Nya sehingga penulis dapat menyelesaikan salah satu kewajiban penulis sebagai mahasiswa Departemen Teknik Informatika ITS, yaitu Kerja Praktik yang berjudul: **PELABELAN DAN PEMBUATAN MODEL IMAGE CAPTIONING MENGGUNAKAN DEEP LEARNING**.

Penulis menyadari bahwa terdapat banyak kekurangan dalam pelaksanaan kerja praktik dan penyusunan buku laporan kerja praktik ini. Namun, penulis berharap buku laporan ini dapat menambah wawasan pembaca dan dapat menjadi sumber referensi untuk mengembangkan keilmuan di masa depan.

Melalui buku laporan ini penulis juga ingin menyampaikan terima kasih kepada orang-orang yang telah membantu menyusun laporan kerja praktik baik secara langsung maupun tidak langsung antara lain:

1. Kedua orang tua penulis.
2. Bapak Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc. selaku dosen pembimbing departemen pada kerja praktik ini.
3. Ibu Dini Adni Navastara, S.Kom, M.Sc. selaku dosen pembimbing lapangan pada kerja praktik ini.
4. Teman-teman penulis yang senantiasa memberikan semangat ketika penulis melaksanakan KP.

Surabaya, 30 Januari 2025
Wardatul Amalia Safitri dan Hammuda Arsyad

[Halaman ini sengaja dikosongkan]

BAB I
PENDAHULUAN

1.1. Latar Belakang

Dalam beberapa tahun terakhir, kemajuan teknologi *deep learning* telah mendorong pengembangan berbagai penerapan berbasis visi komputer, termasuk dalam bidang pengolahan gambar dan pemahaman bahasa alami. Salah satu penerapan *deep learning* yang menarik perhatian adalah pengembangan teknologi *image captioning*. *Image captioning* adalah proses menghasilkan deskripsi teks yang relevan dan bermakna berdasarkan sebuah gambar yang tersedia. *Image captioning* memiliki potensi pemanfaatan yang luas dalam kehidupan, salah satunya untuk membantu penyandang disabilitas visual dalam mengenali lingkungan sekitarnya. Namun, untuk mencapai hasil deskripsi yang optimal, pengembangan model *image captioning* membutuhkan dataset yang dilabeli dengan baik dan metode pembelajaran mesin yang efisien.

Pelabelan data merupakan langkah penting dalam pengembangan model *image captioning* karena deskripsi teks yang akurat dan sesuai konteks sangat memengaruhi performa model. Proses pelabelan ini memerlukan pemahaman mendalam terhadap isi gambar serta kemampuan untuk menerjemahkannya ke dalam kalimat yang bermakna. Tantangan utama dalam pelabelan data adalah memastikan konsistensi deskripsi, terutama ketika dataset terdiri dari ribuan hingga jutaan gambar dengan variasi konteks yang beragam. Oleh karena itu, diperlukan pendekatan sistematis untuk membuat dataset dengan label yang siap digunakan untuk pelatihan model *deep learning*. Proses pelabelan dan

pengembangan model yang telah dilakukan sebelumnya, belum menunjukkan hasil yang memuaskan. Perlu adanya perubahan metode pelabelan dan eksplorasi metode pembelajaran mesin agar hasil *captioning* menunjukkan peningkatan jika dibandingkan dengan metode sebelumnya.

1.2. Tujuan

Tujuan dari kerja praktik ini adalah menerapkan ilmu-ilmu informatika yang didapatkan di bangku perkuliahan ke dalam proses pembuatan model *image captioning*. Secara lebih terperinci, tujuan dari kerja praktik ini adalah:

a) Membuat dataset citra beserta label untuk pengembangan *image captioning*.
b) Membuat model *image captioning* dengan mengimplementasikan berbagai metode *deep learning*.
c) Mengevaluasi metode pelabelan dan model pembelajaran mesin yang digunakan.

1.3. Manfaat

Adapun manfaat pelabelan dan pengembangan model *image captioning* yang dilakukan dalam kerja praktik ini adalah sebagai berikut:

a) Menghasilkan dataset dengan label yang berkualitas dan dapat digunakan untuk pengembangan lainnya.
b) Menghasilkan model *image captioning* yang dapat diimplementasikan dalam berbagai hal yang aplikatif.
c) Mengetahui pengaruh metode pelabelan dan metode pembelajaran mesin terhadap performa model dalam *image captioning*.
d) Menjadi referensi dalam mengembangkan teknologi *image captioning* di masa depan.

1.4. Rumusan Masalah

Berikut ini rumusan masalah pada kerja praktik pelabelan dan pembuatan model *image captioning* menggunakan *deep learning*:

a) Bagaimana cara mengumpulkan citra dan melabeli dataset citra agar bisa menghasilkan model *image captioning* yang berkualitas?
b) Bagaimana metode *deep learning* yang dapat diimplementasikan untuk membuat model *image captioning* yang berkualitas?
c) Bagaimana cara melakuakan evaluasi performa model *image captioning*?

1.5. Lokasi dan Waktu Kerja Praktik

Kerja praktik ini dilaksanakan pada waktu dan tempat sebagai berikut:

Lokasi	: Laboratorium Komputasi Cerdas dan Visi (KCV) Departemen Teknik Informatika ITS
Waktu	: 1 September – 1 Desember 2024
Hari Kerja	: Fleksibel
Jam Kerja	: Fleksibel

1.6. Metodologi Kerja Praktik

1.6.1. Perumusan Masalah

Dalam tahap ini kami perlu mengetahui permasalahan apa saja yang terjadi dan dapat diselesaikan. Kami juga perlu mengetahui semua kebutuhan dalam permasalahan tersebut.

1.6.2. Studi Literatur

Setelah ditentukan rumusan masalah mengenai sistem yang akan dibuat, selanjutnya dilakukan studi literatur mengenai detail implementasinya. Pada tahap ini dilakukan proses pencarian, pembelajaran, dan pengumpulan informasi tentang *image captioning*, metode *deep learning* dalam pemodelan *image captioning*, tahap-tahap yang perlu dilakukan untuk meningkatkan hasil deskripsi oleh sistem, dan metrik evaluasi performa.

1.6.3. Analisis dan Perancangan

Pada tahap ini, dijelaskan hasil dari studi literatur yang telah dilakukan. Dari berbagai metode yang ditemukan selama proses literasi, dianalisis metode yang dapat diterapkan untuk menyelesaikan masalah. Selain itu, ditentukan bahasa pemrograman yang akan digunakan serta batasan data yang akan dimanfaatkan, sehingga hasil yang diharapkan dapat dicapai.

1.6.4. Implementasi Sistem

Pada tahap ini dijelaskan implementasi program yang digunakan pada proses pembuatan model *image captioning.* Sebagai gambaran umum, proses pembuatan model akan menggunakan bahasa pemrograman Python. Pada pembuatan model *image captioning,* dataset yang digunakan berupa dataset citra trotoar dan lingkungan Institut Teknologi Sepuluh Nopember (ITS). Dataset ini akan digunakan pada berbagai skema pemodelan *image captioning*.

1.6.5. Pengujian dan Evaluasi

Setelah proses persiapan data dan pelatihan sistem, perlu dilakukan evaluasi performa model untuk menguji apakah model memiliki kualitas yang baik atau tidak. Pengujian akan memanfaatkan BLEU score untuk membandingkan deskripsi dari model terhadap deskripsi yang telah diberikan pada proses pelabelan sebelumnya.

1.6.6. Kesimpulan dan Saran

Pada bab ini, dipaparkan kesimpulan yang dapat diambil dan juga saran dalam pengerjaan kerja praktik.

1.7. Sistematika Laporan

Laporan kerja praktik ini terdiri dari tujuh bab dengan rincian sebagai berikut:

1.7.1. Bab I Pendahuluan

Bab ini berisi latar belakang, tujuan, manfaat, rumusan masalah, lokasi dan waktu kerja praktik, metodologi, dan sistematika laporan.

1.7.2. Bab II Profil Perusahaan

Bab ini berisi gambaran umum Laboratorium Komputasi Cerdas dan Visi (KCV) Departemen Teknik Informatika ITS mulai dari profil, lokasi perusahaan, dan struktur organisasi.

1.7.3. Bab III Tinjauan Pustaka

Bab ini berisi dasar teori dari teknologi yang digunakan dalam menyelesaikan proyek kerja praktik.

1.7.4. Bab IV Analisis dan Perancangan Infrastruktur Sistem

Bab ini berisi mengenai tahap analisis sistem dalam menyelesaikan proyek kerja praktik.

1.7.5. Bab V Implementasi Sistem

Bab ini berisi uraian tahap - tahap yang dilakukan untuk proses implementasi.

1.7.6. Bab VI Pengujian dan Evaluasi

Bab ini berisi hasil uji coba dan evaluasi dari sistem yang telah dikembangkan selama pelaksanaan kerja praktik.

1.7.7. Bab VII Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran yang didapat dari proses pelaksanaan kerja praktik.

BAB II
PROFIL LOKASI KERJA PRAKTIK

2.1. Logo Lokasi Kerja Praktik

Laboratorium Komputasi Cerdas dan Visi (KCV) Institut Teknologi Sepuluh Nopember (ITS) adalah salah satu lab di Departemen Teknik Informatika ITS yang berfokus pada penelitian dan pengembangan di bidang kecerdasan buatan, visi komputer, dan komputasi cerdas. Laboratorium ini mendukung kegiatan akademik dan riset dengan menyediakan sarana dan prasarana modern untuk mahasiswa dan peneliti, termasuk untuk tugas akhir, proyek penelitian, dan kolaborasi multidisiplin. Dengan berbagai program riset inovatif, lab ini bertujuan untuk memberikan kontribusi signifikan terhadap perkembangan ilmu pengetahuan dan teknologi di Indonesia.

2.2. Logo Lokasi Kerja Praktik

Gambar 2.1 di bawah ini adalah logo Laboratorium Komputasi Cerdas dan Visi (KCV) Teknik Informatika ITS.

Gambar 2.1 Logo Laboratorium KCV

2.3. Alamat Lokasi Kerja Praktik

Jalan Teknik Kimia - Gedung Departemen Teknik Informatika, Kampus Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya, Jawa Timur.

2.4. Anggota Lokasi Kerja Praktik

Daftar anggota lab KCV 2024 adalah sebagai berikut:

a. Kepala Lab: Dr. Eng. Nanik Suciati, S.Kom., M.Kom.
b. Dosen anggota:
   1) Prof. Dr. Eng. Chastine Fatichah, S.Kom., M.Kom.
   2) Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D.
   3) Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom.
   4) Dini Adni Navastara, S.Kom., M.Sc.
   5) Aldinata Rizky Revanda, S.Kom., M.Kom.
   6) Imam Mustafa Kamal, S.ST, Ph.D.
c. Teknisi Lab: Ahmad Junaidi Abdillah, S.Kom.

2.5. Fasilitas Lokasi Kerja Praktik

Laboratorium Komputasi Cerdas dan Visi dilengkapi dengan komputer yang memiliki spesifikasi Processor Intel Core i3 Gen-3, i5 Gen-8, Intel® Xeon® E5-2640 dengan RAM 4GB-16GB, sampai dengan Processor i9 Generasi 12 GPU 3080TI. Kapasitas HDD sebagian besar perangkat setidaknya 1TB. Semua monitor memiliki ukuran 19″ untuk memudahkan mahasiswa dalam melakukan penelitian dan pembelajaran. Laboratorium juga dilengkapi LED TV 55″ serta LCD Projector untuk memunjang demo pembelajaran atau presentasi hasil pekerjaan.

2.6. Penelitian dan Pengabdian

Laboratorium Komputasi Cerdas dan Visi (KCV) telah melakukan berbagai kegiatan penelitian dan pengabdian. Pada tahun 2023, Laboratorium KCV mengadakan kegiatan pengabdian berupa "Pemanfaatan Teknologi Informasi untuk Peningkatan Kualitas Pembelajaran pada SD Yapita Surabaya". Di tahun yang sama, beberapa penelitian yang telah dilakukan adalah:

a. Pengembangan Metode Deteksi Subtipe Acute Lymphoblastic Leukemia (ALL) Pada Citra Mikroskopis Sel Darah Sebagai Alat Bantu Diagnosis Tipe Leukemia.
b. Pengembangan Sistem Automatic Medical Report Generation Berdasarkan Citra Medis.
c. Kerangka Kerja Model Pengembangan Sistem Gamifikasi Adaptif Untuk Meningkatkan Efektivitas Pengerjaan Tugas Akhir Mahasiswa.
d. Sistem Pendeteksian Dan Pemantauan Dini Risiko Komplikasi Berbasis Multimodal Deep Learning Pada Pasien Yang Menjalani Terapi Contiuous Ambulatory Peritoneal Dialysis.
e. Aplikasi Penghitung dan Pengukur Kecepatan Kendaraan Terpadu Berbasis Deep Learning pada Survey Lalu Lintas Harian Rata-Rata.
f. Pengembangan Model Deteksi Ketersediaan Slot Parkir Dari Dua Kamera Yang Overlap Menggunakan Yolo Dan Image Stitching.
g. Deep Learning untuk Deteksi Jenis dan Tingkat Kesegaran Ikan, Penelitian Disertasi Doktor.

h. Pengenalan Pelat Nomor Kendaraan Menggunakan Convolutional Recurrent Neural Network dengan Augmentasi Geometri, Cuaca dan Cahaya.
i. Pengenalan Lingkungan Bagi Tuna Netra Menggunakan Image Captioning Berbasis Deep Learning.

BAB III
TINJAUAN PUSTAKA

3.1. Image Captioning

Image captioning adalah salah satu cabang penting dalam bidang *computer vision* dan *natural language processing* yang bertujuan untuk menghasilkan deskripsi tekstual dari gambar. Proses ini menggabungkan teknik pengolahan gambar untuk mengekstrak fitur visual dengan model bahasa untuk menyusun deskripsi yang sesuai dalam bentuk kalimat. Teknologi ini banyak diterapkan dalam berbagai domain, seperti aplikasi pembantu bagi penyandang disabilitas penglihatan, pencarian gambar berbasis teks, dan analisis konten media sosial. Pendekatan modern dalam *image captioning* umumnya menggunakan arsitektur berbasis *deep learning*, seperti kombinasi *Convolutional Neural Networks* (CNN) untuk ekstraksi fitur visual dan *Recurrent Neural Networks* (RNN), *Long Short-Term Memory* (LSTM), atau transformer untuk generasi teks. Penelitian di bidang ini terus berkembang dengan fokus pada peningkatan akurasi, efisiensi, dan pemahaman konteks yang lebih mendalam antara elemen visual dan linguistik (Ghandi et al., 2022).

Gambar 3.1 Contoh Gambar untuk Image Captioning

Gambar 3.1 menunjukkan sebuah ruang kelas yang sedang digunakan untuk pembelajaran. Ada banyak orang, yang kemungkinan adalah mahasiswa, sedang duduk di kursi yang disediakan. Ada juga seseorang yang berdiri di depan ruang kelas. Di langit-langit ruang kelas terdapat banyak lampu dan sebuah proyektor. Dari sebuah gambar ini terdapat banyak hal yang bisa dijelaskan. Oleh karena itu, ketika Gambar 3.1 dimasukkan ke dalam sebuah sistem *image captioning*, ada banyak kemungkinan deskripsi gambar yang bisa diberikan oleh sistem. Deskripsi terbaik adalah deskripsi yang memiliki fokus pembahasan sesuai dengan konteks utama yang ditunjukkan pada gambar. Misalnya, "Mahasiswa memenuhi ruangan kelas, sebagian memperhatikan seorang presenter yang sedang berdiri di depan dekat proyektor".

3.2. Deep Learning

Deep learning adalah salah satu pendekatan dalam pembelajaran mesin yang menggunakan jaringan saraf tiruan dengan banyak lapisan (*deep neural networks*) untuk memproses dan memahami data dalam berbagai bentuk. Metode ini mampu secara otomatis mengekstrak fitur kompleks dari data mentah, sehingga sangat efektif dalam menangani masalah dengan dimensi data yang tinggi, seperti pengolahan citra, pengenalan suara, dan pemrosesan bahasa alami. Keunggulan *deep learning* terletak pada kemampuannya untuk mempelajari representasi data yang lebih abstrak dan bermakna dibandingkan metode pembelajaran tradisional. Model *deep learning*, seperti

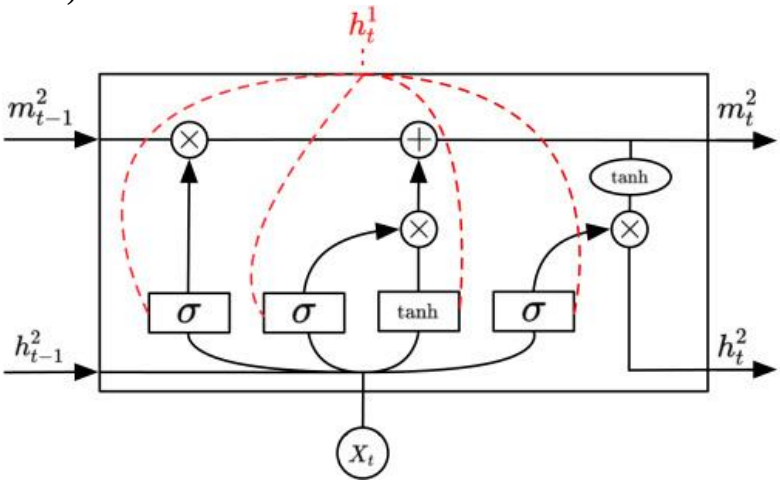
Convolutional Neural Networks (CNN) untuk analisis citra dan *Recurrent Neural Networks* (RNN) atau transformer untuk data berurutan, telah menjadi standar dalam banyak aplikasi. Kemajuan dalam komputasi, seperti penggunaan GPU dan TPU, serta ketersediaan dataset besar, turut mendukung perkembangan pesat *deep learning* dalam berbagai domain penelitian dan industri (Castro et al., 2022).

3.3. LSTM

Pada pembelajaran mesin (*machine learning*), model LSTM banyak digunakan dalam masalah analisis sekuensial. LSTM memiliki dua properti utama, yaitu *context sensitive learning* dan *good generalization*. Arsitektur LSTM sering kali dikenal sebagai pengenal teks yang generik dan tidak tergantung bahasa. Oleh karena itu, LSTM telah banyak digunakan untuk mengenali teks tulisan tangan mauouan tulisan cetak dengan sukses. Pada LSTM, terdapat sebuah mekanisme yang memungkinkan sebuah sel pada LSTM memiliki kemampuan untuk menghapus informasi di dalam sel yang sudah tidak dibutuhkan dan menambahkan informasi baru ke dalam sel.

Adapun metode yang lebih kompleks, yakni Bidirectional *Long Short-Term Memory* (BiLSTM), memiliki dua jaringan LSTM dimana jaringan LSTM pertama berfungsi dalam memproses urutan input data ke arah depan (*forward*) dan jaringan LSTM kedua berfungsi dalam memproses urutan data dari arah sebaliknya (*backward*) (Puteri, 2023). Kemudian output dari jaringan LSTM *forward* dan *backward* digabungkan pada setiap urutan waktu. Dengan adanya dua lapisan yang berlawanan arah tersebut, model dapat mempelajari informasi masa lalu dan informasi

masa mendatang untuk setiap *sequence input* (Kesiman et al., 2021).

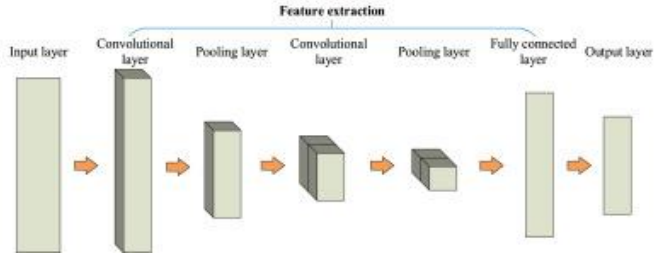


Gambar 3.2 Arsitektur LSTM (J. Zhang et al., 2021)

3.4. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah model *deep learning* yang paling banyak digunakan dalam pembelajaran fitur untuk klasifikasi dan pengenalan gambar skala besar. Lapisan konvolusional menggunakan operasi konvolusi untuk mencapai pembagian bobot, sedangkan lapisan subsampling digunakan untuk mengurangi dimensi. Lapisan subsampling bertujuan untuk mengurangi dimensi fitur. Lapisan subsampling biasanya dapat diimplementasikan dengan operasi *average pooling* atau operasi *max pooling*. Setelah itu, beberapa lapisan yang terhubung penuh dan lapisan softmax biasanya diletakkan di lapisan atas untuk klasifikasi dan pengenalan objek (Q. Zhang et al., 2018).

Deep convolutional neural network biasanya mencakup beberapa lapisan konvolusional dan lapisan subsampling untuk pembelajaran fitur pada gambar skala besar. Dalam beberapa tahun terakhir, CNN juga meraih kesuksesan besar dalam pemrosesan bahasa, pengenalan suara, dan sebagainya. Salah satu penerapan CNN dalam image captioning dilakukan oleh Ruifan Li, dkk. Dalam penelitiannya, dual-CNN diterapkan pada decoder untuk membuat model *image captioning* (R. Li et al., 2020).

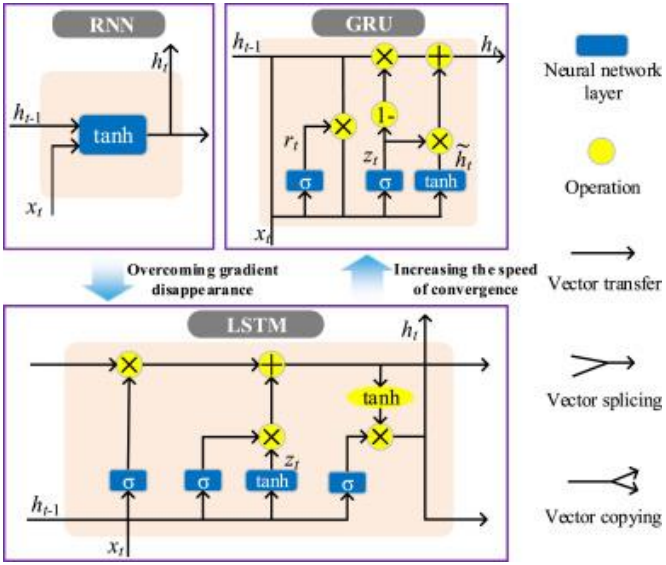


Gambar 3.3 Arsitektur CNN (C. Li et al., 2022)

3.5. Gated Recurrent unit (GRU)

Gated Recurrent Unit (GRU) diusulkan oleh Cho et al. (2014) sebagai salah satu RNN yang sebanding dengan LSTM. GRU memiliki arsitektur yang lebih sederhana daripada LSTM dan menggunakan *update and reset gates* untuk menangani aliran informasi di *hidden unit* (de Rautlin de la Roy et al., 2023). Saat memproses *long-time sequence*, Recurrent Neural Network (RNN) mengalami fenomena kehilangan gradien. Oleh karena itu, para peneliti telah meningkatkan RNN dengan menambahkan unit sel dan tiga gerbang kontrol di lapisan tersembunyi untuk membuat jaringan LSTM. LSTM cocok untuk memprediksi data *time-series*, namun struktur LSTM cukup kompleks sehingga

waktu yang digunakan untuk pelatihan data cukup lama. Namun, GRU menyederhanakan struktur LSTM dengan menggabungkan *input gates* dan *forgetting gates* menjadi *update gates*, yang mempercepat konvergensi selama pelatihan. RNN ditingkatkan dari LSTM ke GRU, dan perubahan strukturalnya ditunjukkan pada Gambar 3.4.



Gambar 3.4 Perbandingan Arsitektur RNN, LSTM, dan GRU (C. Li et al., 2022)

3.6. BLEU Score

Bilingual Evaluation Understudy (BLEU) merupakan algoritma untuk membandingkan frasa berturut-turut dari terjemahan otomatis dengan frasa berturut-turut yang ditemukannya dalam terjemahan referensi, dan menghitung jumlah kecocokan, dengan menggunakan konstanta yang dinamakan *brevity penalty*. Proses pencocokan

menggunakan BLEU Score dilakukan secara independen. Tingkat kecocokan yang lebih tinggi menunjukkan tingkat kesamaan yang lebih tinggi terhadap terjemahan referensi. Pada pencocokan menggunakan BLEU Score, kejelasan dan ketepatan tata bahasa tidak diperhitungkan. Kecocokan sempurna akan menghasilkan skor 1.0, sedangkan ketidakcocokan sempurna menghasilkan skor 0.0. Nilai BLEU dapat dihitung dengan Persamaan 2.1-2.3.

$$Unigram\ Precision\ P = \frac{m}{W_t} \tag{2.1}$$

$$Brevity\ penalty\ p = \begin{cases} 1\ jika\ c\ > r \\ e^{\left(1-\frac{r}{c}\right)}\ jika\ c\ \leq r \end{cases} \tag{2.2}$$

$$BLEU = \ p\ .\ e^{\sum_{n=1}^{N}\left(\frac{1}{N}logP_n\right)} \tag{2.3}$$

Dimana m adalah jumlah kata dari kandidat yang ditemukan dalam referensi. Wt adalah total jumlah kata pada kandidat. r adalah panjang efektif korpus acuan, dan c total panjang korpus terjemahan. Pn adalah rata-rata geometris dari n-gram yang dimodifikasi presisi. N adalah panjang n-gram yang digunakan untuk menghitung Pn (Kumari et al., 2011).

BLEU Score terdiri dari empat jenis, yaitu BLEU-1, BLEU-2, BLEU-3, dan BLEU-4. Dalam konsep BLEU, terdapat istilah gram yang dapat diartikan sebagai jumlah kata. BLEU-4 didapatkan dengan menghitung nilai kecocokan dari 1-gram hingga 4-gram, dimana masing masing gram memiliki bobot yang sama yaitu 0,25. Contoh pada kalimat yang dihasilkan suatu model yaitu “saya sedang makan nasi di meja”. Maka kalimat tersebut akan dibagi setiap empat kata yaitu “saya sedang makan nasi”, “sedang

makan nasi di", dan "makan nasi di meja". Dari ketiga kemungkinan potongan kalimat tersebut, dilakukan pencocokan terhadap kalimat asli yang menjadi referensi. Adapun kalimat referensi ini juga dipisah setiap empat kata dan dihitung jumlah kemungkinan yang benar. Nilai BLEU di atas 30 mencerminkan kalimat yang dapat dimengerti. Sedangkan nilai BLEU di atas 50 mencerminkan kalimat yang baik dan fasih (Lavie, 2011).

3.7. ROUGE Score

Recall-Oriented Understudy for Gisting Evaluation **(**ROUGE) adalah metrik evaluasi yang menghitung unit yang tumpang tindih antara kalimat yang dihasilkan sebuah sistem terhadap kalimat referensi Terdapat beberapa konsep ROUGE untuk evaluasi performa, diantaranya ROUGE-N dan ROUGE-L. ROUGE-N mengukur n-gram yang tumpang tindih antara kalimat yang dihasilkan sebuah sistem terhadap kalimat referensi. Jika N=1 maka nilai yang dihitung adalah banyaknya unigram yang sama antara kedua kalimat. Jika N=2 maka nilai yang dihitung adalah banyaknya bigram yang sama antara kedua kalimat. Begitu pun dengan nilai n yang lebih tinggi (Sanchez-Gomez et al., 2022). Formula untuk menghitung ROUGE-N dapat dilihat pada persamaan 2.4.

$$ROUGE - N = \frac{Number\ of\ matching\ n{-}grams}{Total\ n{-}grams\ in\ the\ reference} \qquad (2.4)$$

ROUGE-L didasarkan pada panjang *Longest Common Subsequence* (LCS). Ini menghitung rata-rata harmonik berbobot (*f-measure*) yang menggabungkan *precision score* dan *recall score*. Persamaan untuk menghitung ROUGE-L adalah persamaan 2.5.

$$ROUGE - L = \frac{(1+\beta^2).P.R}{\beta^2.P+R} \quad (2.5)$$

Dimana P adalah Precision, R adalah recall, dan adalah sebuah konstanta yang biasanya diset dengan nilai 1.

[Halaman ini sengaja dikosongkan]

BAB IV
ANALISIS DAN PERANCANGAN INFRASTRUKTUR SISTEM

4.1. Analisis Sistem

Pada bab ini akan dijelaskan mengenai tahapan dalam pelabelan dan pembuatan model *image captioning* menggunakan *deep learning*. Hal tersebut akan dijelaskan ke dalam dua bagian, yaitu definisi umum dan analisis kebutuhan sistem.

4.1.1. Definisi Umum Pengembangan Sistem

Secara umum, tahap pengerjaan kerja praktik ini dibagi menjadi tahap persiapan dataset dan pembuatan model. Pada tahap persiapan dataset, setiap citra yang ada dalam dataset akan diberi kalimat deskripsi. Citra yang digunakan diambil dari dataset trotoar (Navastara et al., 2023) dan citra baru yang diambil di lingkungan Institut Teknologi Sepuluh Nopember (ITS). Pembuatan kalimat deskripsi akan dilakukan oleh 3 orang secara independen. Satu kalimat deskripsi juga didapatkan dari hasil *generate AI*. Serta, tambahan 2 kalimat deskripsi yang didapatkan dari proses pelabelan yang telah dilakukan pada pengembangan sebelumnya. Dengan demikian, setiap gambar akan memiliki beberapa kalimat deskripsi atau *caption* dari sumber yang beragam.

Tahap berikutnya adalah proses pembuatan model *image captioning*. Proses pemodelan dilakukan menggunakan bahasa Python. Tahap pembuatan model meliputi *import* dataset, *pre-process* data, *data training*, *data testing*, dan evaluasi performa model. Terdapat beberapa skenario pemanfaatan *deep learning* dalam pembuatan model

ini. Sedangkan metode pengukuran evaluasi performa model akan dilakukan menggunakan BLEU Score dan ROUGE.

4.1.2. Identifikasi Kebutuhan Sistem

Untuk melakukan seluruh rangkaian tahapan kerja praktik ini, terdapat beberapa kebutuhan baik software maupun hardware yang diperlukan. Kebutuhan-kebutuhan tersebut meliputi:

1. Software

a. Excel, digunakan untuk pelabelan dataset.
b. Python, bahasa pemrograman yang digunakan untuk pembuatan model.
c. Kaggle, platform berbasis cloud yang digunakan sebagai pusat penyimpanan dataset dan model dalam pengerjaan kerja praktik ini.
d. Google Collab, platform kolaborasi pembuatan model training pipeline.
e. Visual Studio Code, aplikasi yang digunakan untuk eksekusi kode program pelatihan model.
f. Tensorflow, library utama yang digunakan untuk membuat model *image captioning*.

2. Hardware

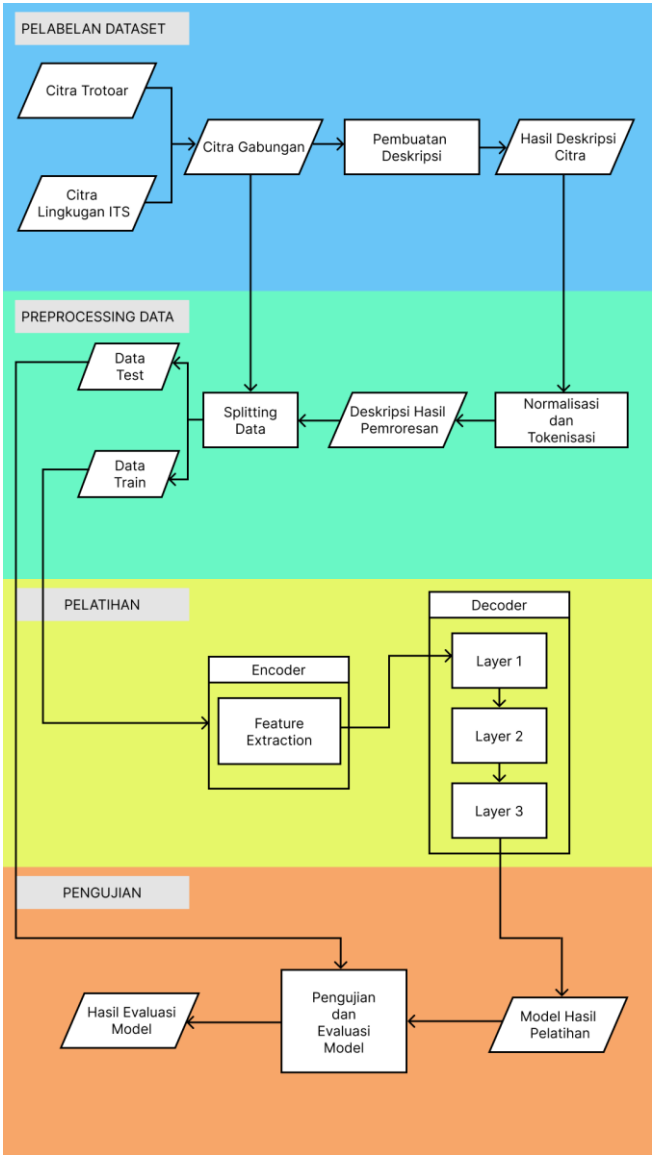
a. Komputer dengan spesifikasi:
i. CPU : Intel Core I9-13900KS
ii. GPU : NVIDIA Geforce RTX 4080 SUPER
iii. Memory : 64 GB

b. Handphone, digunakan untuk mengambil dataset gambar baru.

4.2. Perancangan Infrastruktur Sistem

4.2.1. Desain Sistem

Sistem yang akan dibuat dapat dikelompokkan menjadi 4 tahap, yaitu pelabelan dataset, preprocessing data, pelatihan (*training*), dan pengujian (*testing*). Diagram alir dari sistem yang akan dibuat dalam kerja praktik ini dapat dijelaskan pada Gambar 4.1. Pada tahap pelatihan model, terdapat beberapa skenario yang akan diimplementasikan dan dibandingkan performanya. Setiap skenario akan mengimplementasikan metode yang berbeda pada lapisan-lapisan decoder. Skenario yang akan diimplementasikan meliputi penerapan LSTM, CNN, dan GRU.



Gambar 4.1 Diagram Alir Rancangan Sistem

4.2.2. Pelabelan Dataset

Citra yang digunakan pada kerja praktik ini sebagian diambil dari dataset trotoar (Navastara et al., 2023) yang selanjutnya disebut sebagai Dataset 1. Dataset 1 dibuat dengan cara mengambil video pada beberapa titik trotoar di Surabaya dan Gresik. Lalu data video tersebut diambil dan diseleksi berdasarkan frame yang memiliki tingkat blur rendah dengan menggunakan metode laplacian. Terdapat 1.447 citra pada Dataset 1 yang akan digunakan pada penelitian ini. Metode pelabelan yang digunakan pada Dataset 1 terdiri dari 5 kalimat deskripsi untuk setipa citra. Kelima kalimat ini dideskripsikan oleh satu orang saja. Selain itu, dataset 1 mencantumkan dataset 1693 citra dari dataset flickr30k.

Pada kerja praktik ini akan dilakukan pelabelan ulang pada citra Dtaaset 1 khusus citra trotoar. Pelabelan akan dilakukan oleh 3 orang secara independen. Kemudian, akan ditambahkan 1 label dari hasil *generate AI* dan tambahan 2 kalimat dari Dataset 1. Sehingga pada Dataset 2 setiap citra memiliki 6 kalimat deskripsi dari sumber yang beragam.

Selain itu, pada kerja praktik ini akan dibuat dataset dengan citra baru yang selanjutnya akan digabungkan dengan citra dan label Dataset 2. Dataset gabungan ini akan disebut sebagai Dataset 3. Terdapat 1.554 citra baru yang diambil dari objek-objek yang ada di lingkungan Institut Teknologi Sepuluh Nopember (ITS). Lokasi pengambilan citra adalah Departemen Teknik Informatika, pusat robotika, perumahan dosen, asrama mahasiswa, K1 Mart, Co-Working Space, fasilitas olahraga, ATM, masjid, perpustakaan, Rektorat, kantin, dan tempat parkir di lingkungan pusat ITS.

Untuk setiap citra tambahan pada Dataset 3, metode pelabelan dilakukan dengan cara yang serupa dengan

pelabelan Dataset 2. Terdapat 3 orang yang akan melakukan pelabelan secara manual dan independen. Selain itu, 1 kalimat deskripsi tambahan akan dibuat dengan memanfaatkan kecerdasan buatan atau *Artificial Intelligence* (AI). Dengan demikian, citra pada Dataset 3 sebagian memiliki 6 kalimat deskripi, sedangkan sebagian lainnya memiliki 4 kalimat deskripsi. Setiap kalimat deskripsi yang dibuat minimal memuat 8 kata.

4.2.3. Preprocessing Data

Tahap *preprocessing* atau pra proses merupakan tahapan untuk mengubah teks deskripsi dalam bentuk kalimat ke dalam bentuk token sehingga dapat diolah menjadi model. Setiap kalimat deskripsi yang telah dibuat pada tahap pelabelan dataset akan dilakukan normalisasi pada teks sehingga teks menjadi seragam dan hanya diambil bagian-bagian yang diperlukan saja untuk diolah pada tahap selanjutnya. Penyesuaian yang dilakukan antara lain mengubah teks menjadi *lowercase*, menghilangkan karakter yang tidak diperlukan, seperti angka, spasi ekstra, dan juga menambahkan ‘startseq’ serta ‘endseq’ pada awal dan akhir teks.

Selanjutnya, setiap kalimat akan dipecah menjadi kata-kata penyusunnya yang biasa disebut token. Setiap token diberikan suatu indeks yang berbeda beda. Indeks tersebut urut berdasarkan kata yang muncul terlebih dahulu pada deskripsi dari dataset. Pada akhir *preprocessing data*, citra beserta deskripsi hasil normalisasi dan tokenisasi akan dipecah menjadi data untuk pelatihan dan data untuk pengujian.

4.2.4. Data Training

Tahap ini memiliki tujuan untuk melatih model menggunakan data deskripsi dan citra sehingga model dapat menghasilkan kata demi kata dan membentuk suatu kalimat yang utuh berdasarkan citra. Proses pelatihan dilakukan untuk setiap satu pasangan citra dan kalimat deskripsi.

Pada tahap pertama, proses training dimulai pada bagian encoder. Pada bagian ini, dilakukan proses ekstraksi fitur dengan tujuan untuk mendapatkan fitur baru dari citra. Fitur baru yang dihasilkan berupa objek-objek yang terdapat pada sebuah citra. Proses ekstraksi fitur memanfaatkan arsitektur DenseNet-201 dengan menghilangkan layer terakhir yang berfungsi sebagai labelling. Kemudian fitur-fitur tersebut akan diteruskan ke bagian decoder.

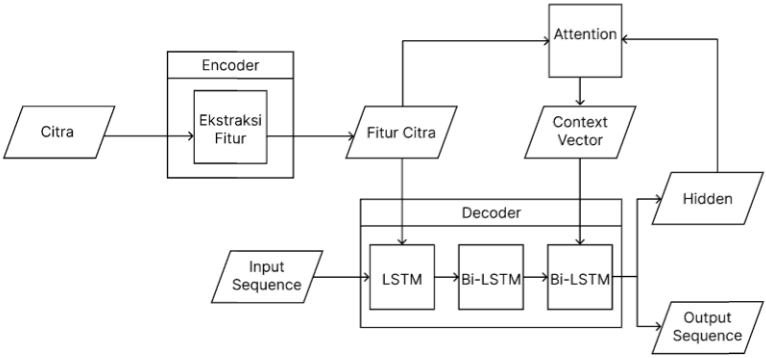
Bagian decoder secara umum akan menghasilkan kata-kata yang nantinya dapat disusun menjadi sebuah kalimat deskripsi. Bagian decoder terdiri dari 3 lapisan pembelajaran mesin untuk mengolah citra dan deskripsi. Input untuk lapisan pertama decoder adalah fitur citra yang didapatkan dari proses encoding serta sequence deskripsi setiap citra. Hasil dari lapisan pertama akan digunakan sebagai input lapisan kedua. Kemudian, hasil dari lapisan kedua akan dijadikan input lapisan ketiga. Hasil dari lapisan ketiga adalah *hidden state* dan *sequence* deskripsi output.

Hidden state dari setiap iterasi decoder akan diolah oleh fungsi *attention* untuk menghasilkan sebuah *context vector*. Fungsi *attention* akan memanfaatkan fitur citra hasil ekstraksi oleh encoder untuk membantu menghasilkan *context vector*. Context *vector* akan dijadikan sebagai input tambahan pada layer ketiga decoder. Penambahan *context vector* sebagai input, dapat membantu mesin menghubungkan konteks sebuah citra pada *timestep* tertentu

terhadap *timestep* sebelumnya. *Context vector* dapat membantu mennetukan fokus deksripsi sebuah citra.

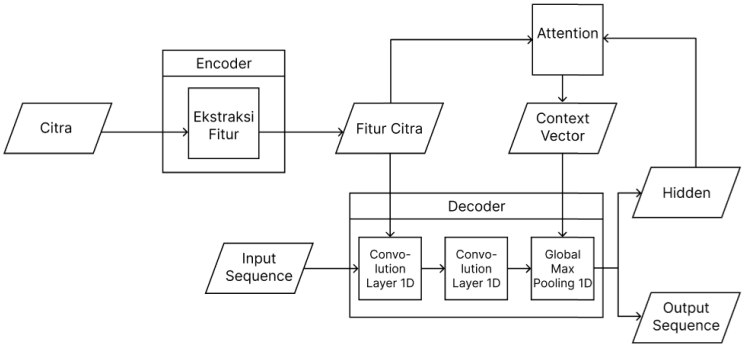
Terdapat 3 skenario pengembangan model *image captioning* pada kerja praktik ini. Setiap skenario akan menerapkan metode *deep learning* berbeda pada lapisan-lapisan decodernya. Skenario yang akan digunakan dalam proses pelatihan data meliputi implementasi LSTM, CNN, dan GRU. Penjelasan lebih detail dari setiap skenario dapat dijelaskan pada Gambar 4.2 - 4.4.

Pada skenario 1, metode *deep learning* yang diterapkan adalah LSTM. Bagian decoder terdiri dari satu lapisan LSTM dan dua lapisan Bi-LSTM.



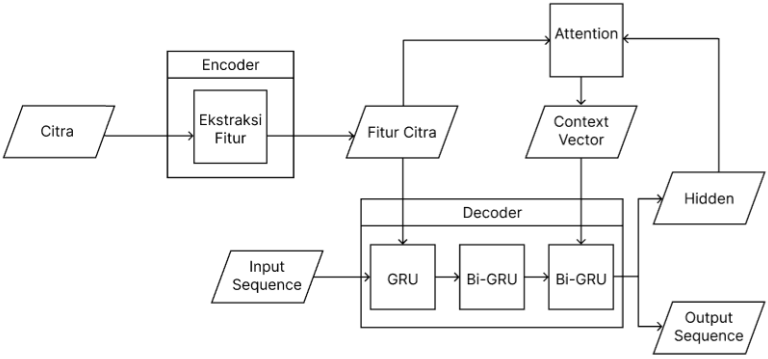
Gambar 4.2 Diagram Alir Skenario Pemodelan 1

Pada skenario 2, metode *deep learning* yang diterapkan adalah CNN. Lapisan pertama dan kedua pada decoder berupa convolution layer 1D. Sedangkan lapisan ketiga berupa global max pooling 1D.



Gambar 4.3 Diagram Alir Skenario Pemodelan 2

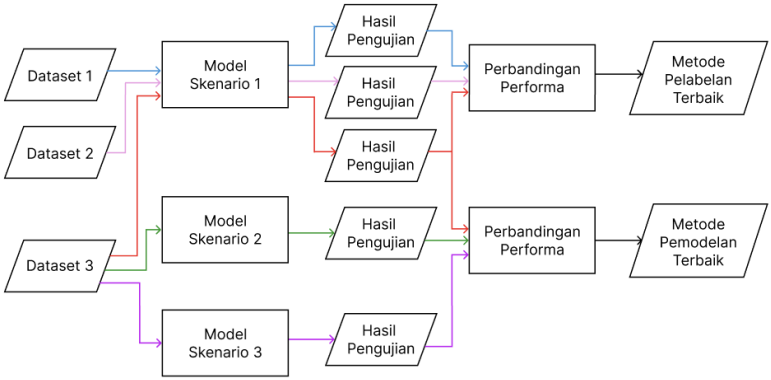
Pada skenario 3, metode yang diterapkan adalah GRU. Pada skenario ini lapisan decoder pertama adalah GRU, sedangkan lapisan decoder kedua dan ketiga adalah Bi-GRU.



Gambar 4.4 Diagram Alir Skenario Pemodelan 3

4.2.5. Data Testing

Pada proses pengujian atau *testing*, citra yang disediakan untuk bahan uji akan dideskripsikan oleh model ke dalam satu kalimat. Kalimat hasil deskripsi oleh model ini akan dibandingkan dengan kalimat deskripsi hasil pelabelan data. Proses perbandingan hasil deskripsi oleh model dengan deskripsi pada proses pelabelan data akan memanfaatkan BLEU Score dan ROUGE sebagai parameter evaluasi performa model. Alur pengujian dan evaluasi performa model dapat dijelaskan melalui diagram alir pada Gambar 4.5.



Gambar 4.5 Diagram Alir Skema Pengujian dan Evaluasi

BAB V
IMPLEMENTASI SISTEM

Bab ini membahas tentang implementasi dari sistem yang telah dibuat untuk pelabelan data dan pemodelan *image captioning.*

5.1. Pelabelan Data

Proses pelabelan data dilakukan menggunakan Microsoft Excel sehingga hasil akhir pelabelan akan disimpan dalam format file csv. Dataset 1 terdiri dari 1.447 citra trotoar yang memiliki 5 kalimat deskripsi pada setiap citranya. Dataset 2 terdiri dari 1.554 citra yang setiap citranya memiliki 4 kalimat deskripsi. Kemudian, Dataset 3 terdiri dari 3.001 citra yang sebagian dari citra tersebut memiliki 4 kalimat deskripsi sedangkan sebagian lainnya memiliki 6 kalimat deskripsi. Berikut adalah contoh deskripsi yang telah dihasilkan pada proses pelabelan dataset.

Tabel 5.1 Contoh Deskripsi pada Dataset 1

No	Citra	Deskripsi
1.		a) trotoar yang cukup sempit ini memiliki 9 buah kubus kecil yang tersebar di dalamnya b) terdapat tempat parkir khusus sepeda motor di area sekitar trotoar

		c) sebuah mobil dengan warna abu-abu dan dua mobil dengan warna hitam terparkir di area sekitar trotoar d) trotoar ini memiliki pilar-pilar besar dengan warna coklat dan hitam e) terdapat banyak tempat duduk berbentuk kotak dan dengan warna abu-abu
2.		a) Terdapat seorang pria dan seorang wanita sedang berjalan b) Beberapa tempat sampah dengan warna merah dan kuning serta dua orang yang sedang berjalan c) Terdapat satu tiang penanda parkir dan dua orang yang sedang berjalan d) Terdapat tiga pohon dan tanaman hias dengan warna hijau yang tumbuh di sekitar pohon e) Terdapat satu tiang parkir sepeda dengan warna abu-abu dan dua orang yang sedang berjalan

3.		a) trotoar ini memiliki dua buah bangku kaleng dengan warna hitam b) di sepanjang trotoar ini terdapat lantai pemandu dengan warna hitam yang terkadang dilalui oleh besi penutup suatu galian c) trotoar ini memiliki beberapa pohon dan beberapa tiang d) trotoar ini memiliki dua buah bangku merah dengan model yang berbeda e) terdapat sebuah golongan kabel dengan warna krim dan seorang pria bertopi putih yang sedang duduk pada kursi
4.		a) Terdapat dua orang laki - laki dengan baju biru sedang berjalan b) Terdapat gerobak yang isinya jualan es krim dan dua orang yang sedang berjalan c) Terdapat satu orang pria yang memakai baju hitam sedang berdiri d) Terdapat satu palang

		penanda dilarang berhenti dan beberapa motor yang terparkir e) Seorang anak Sekolah Dasar yang memakai seragam putih merah sedang duduk dan dua orang laki-laki yang sedang berjalan
5.		a) Dua orang perempuan sedang berjalan bersama di trotoar ini b) Sepanjang trotoar ini dibatasi oleh pilar dan pagar dengan warna hitam c) Terdapat dua pohon yang tumbuh di trotoar ini d) Tanaman hias dengan warna hijau tumbuh diselingi dengan dua pohon e) Terdapat tiga bollard dengan warna oranye putih di trotoar ini

Tabel 5.2 Contoh Deskripsi pada Dataset 2

No	Citra	Deskripsi
1.		a) trotoar yang cukup sempit ini memiliki 9 buah kubus kecil yang tersebar di dalamnya b) terdapat tempat parkir khusus sepeda motor di area sekitar trotoar c) tiang bangunan besar di depan dan beberapa kendaraan parkir di jalan sebelah kanan trotoar d) trotoar yang berisikan beberapa motor dan mobil yang sedang parkir e) Tiang bangunan berkeramik dengan banyak sepeda motor dan mobil terparkir di sekitarnya f) Trotoar di sepanjang jalan dengan deretan mobil yang terparkir dan gedung pencakar langit di latar belakang, menambah nuansa kota modern yang sibuk

2.		a) Terdapat seorang pria dan seorang wanita sedang berjalan b) Beberapa tempat sampah dengan warna merah dan kuning serta dua orang yang sedang berjalan c) Terdapat satu tiang penanda parkir dan dua orang yang sedang berjalan d) Dua pejalan kaki melintasi trotoar, dikelilingi kendaraan, lampu lalu lintas, dan nuansa kota yang hidup e) Di ujung trotoar ada pohon, tempat sampah, dan beberapa orang berjalan menuju jalur zebracross f) Sekelompok orang berjalan di trotoar dekat tempat sampah berwarna kuning dan merah, sementara kendaraan melintas di jalan dengan lampu lalu lintas yang menyala hijau di depan mereka

3.		a) trotoar ini memiliki dua buah bangku kaleng dengan warna hitam b) di sepanjang trotoar ini terdapat lantai pemandu dengan warna hitam yang terkadang dilalui oleh besi penutup suatu galian c) trotoar ini memiliki beberapa pohon dan beberapa tiang d) Trotoar berpola dengan pepohonan, mobil terparkir, dan aktivitas orang-orang di sekitar toko. e) Di tepi trotoar ada 2 mobil parkir berjajar di bawah pohon f) Sebuah trotoar dengan pola ubin berwarna yang menarik, dikelilingi oleh kendaraan terparkir, termasuk mobil hitam dan biru, serta aktivitas pejalan kaki di dekat toko.
4.		a) Terdapat dua orang laki - laki dengan baju biru sedang berjalan b) Terdapat gerobak yang isinya jualan es krim dan

		dua orang yang sedang berjalan c) Terdapat satu orang pria yang memakai baju hitam sedang berdiri d) Menelusuri jalanan yang penuh warna, di mana setiap langkah membawa cerita dan kebersamaan! e) Ada beberapa pria berdiri di tengah trotoar yang di sisi kirinya terdapat pedagang es krim f) Sekelompok anak muda mengenakan kaos biru berjalan di trotoar yang ramai, sementara penjual makanan dan kendaraan terparkir di dekat jalan raya
5.		a) Dua orang perempuan sedang berjalan bersama di trotoar ini b) Sepanjang trotoar ini dibatasi oleh pilar dan pagar dengan warna hitam c) Terdapat dua pohon yang tumbuh di trotoar ini d) Dua wanita berjalan di trotoar, satu membawa payung, dikelilingi

		pepohonan dan suasana kota yang cerah. e) 2 orang di bawah payung berjalan di tengah trotoar f) Dua wanita berjalan di trotoar yang teduh dengan payung, menikmati suasana di bawah pepohonan.

Tabel 5.3 Contoh Deskripsi pada Dataset 3

No	Citra	Deskripsi
1.		a) trotoar yang cukup sempit ini memiliki 9 buah kubus kecil yang tersebar di dalamnya b) terdapat tempat parkir khusus sepeda motor di area sekitar trotoar c) tiang bangunan besar di depan dan beberapa kendaraan parkir di jalan sebelah kanan trotoar d) trotoar yang berisikan beberapa motor dan mobil yang sedang parkir e) Tiang bangunan berkeramik dengan banyak sepeda motor dan

		mobil terparkir di sekitarnya f) Trotoar di sepanjang jalan dengan deretan mobil yang terparkir dan gedung pencakar langit di latar belakang, menambah nuansa kota modern yang sibuk
2.		a) Terdapat seorang pria dan seorang wanita sedang berjalan b) Beberapa tempat sampah dengan warna merah dan kuning serta dua orang yang sedang berjalan c) Terdapat satu tiang penanda parkir dan dua orang yang sedang berjalan d) Dua pejalan kaki melintasi trotoar, dikelilingi kendaraan, lampu lalu lintas, dan nuansa kota yang hidup e) Di ujung trotoar ada pohon, tempat sampah, dan beberapa orang berjalan menuju jalur zebracross f) Sekelompok orang

		berjalan di trotoar dekat tempat sampah berwarna kuning dan merah, sementara kendaraan melintas di jalan dengan lampu lalu lintas yang menyala hijau di depan mereka
3.		a) Suasana tenang di sekitar masjid, terdapat larangan berjualan dan tanaman hijau rimbun. b) Suasana tenang di sekitar masjid ini dilengkapi dengan larangan berjualan dan lingkungan yang nyaman. c) Dikelilingi tanaman hijau yang rimbun, suasana tenang di sekitar masjid ini ditandai dengan larangan berjualan yang jelas. d) Di sekitar masjid, suasana tenang dan asri tercipta dengan larangan berjualan dan pepohonan hijau yang rimbun.

4.		a) Bangunan pos jaga dengan tenda biru, dikelilingi pepohonan dan area olahraga yang tenang. b) Bangunan pos jaga dengan tenda biru yang dikelilingi pepohonan c) Dikelilingi oleh pepohonan dan area olahraga, bangunan pos jaga modern dengan tenda biru menawarkan suasana yang nyaman. d) Suasana tenang di sekitar bangunan pos jaga modern dengan tenda biru dan area olahraga yang asri.
5.		a) Lorong lantai atas sisi kirinya terdapat bangku kayu di dekat dinding yang menempel di dinding sedangkan sisi kanannya adalah dinding pembatas sebagai pengaman b) Lorong lantai atas dilengkapi bangku kayu yang nyaman di sisi kiri untuk bersantai. c) Dinding pembatas di sisi kanan memberikan keamanan tambahan di

		lorong ini. d) Ada bangku di lorong dengan lantai oranye dan pembatas di sisi kanan

5.2. Preprocessing Data

Setelah memasukkan dataset citra dan hasil pelabelan dataset ke dalam sistem pemodelan, nama file citra dan keempat deskripsi dari citra tersebut akan disimpan dalam sebuah dictionary. Selanjutnya, setiap kalimat deskripsi akan dinormalisasi dan ditokenisasi. Algoritma normalisasi dan tokenisasi kalimat deskripsi adalah sebagai berikut.

```
INPUT:
dataset (List of Strings): Daftar caption yang diproses.

OUTPUT:
caption (String): Caption yang telah diproses dan
disiapkan untuk tahap selanjutnya.

1 FOR caption in dataset :
2    convertLowercase(caption)
3    removeSpecialChar(caption)
4    removeOneChar(caption)
5    getOnlyAlphabet(caption)
6    addStartseqEndseq(caption)
7    WRITE caption
```

Pseudocode 5.1 Normalisasi

```
INPUT:
    - dataset (List of Strings): Daftar caption yang
akan diproses.

OUTPUT:
    - InputSeq (Array): Urutan input yang dipadatkan
(padded sequence) dengan panjang yang seragam.

1 Dataset = Tokenize(dataset(caption))
2 InputSeq(32, 1) = PadTextToSameLength(Dataset)
```

Pseudocode 5.2 Tokenisasi

Setelah proses normalisasi dan tokenisasi selesai, dataset akan dipecah menjadi data pelatihan dan data pengujian.

5.3. Data Training

Proses training akan melibatkan proses encoding atau ekstraksi fitur, decoding, penentuan visual attention dan context vector seperti yang telah dijelaskan pada subbab 4.2.4. Untuk proses ekstraksi fitur oleh encoder, algoritma yang digunakan adalah sebagai berikut.

```
INPUT: - embedding_dim (Integer): Dimensi ruang
embedding.
       - x (Tensor): Tensor masukan yang mewakili fitur
gambar.

OUTPUT: - x (Tensor): Tensor yang telah ditransformasi
setelah melewati lapisan Dense dan aktivasi ReLU.
```

```
1 Class Encoder:
2    Initialize(embedding_dim):
3       DenseLayer = Dense(embedding_dim)
4
5 Call(x):
6    x = ReLU(DenseLayer(x))
7    Return x
```

Pseudocode 5.3 Ekstraksi Fitur oleh Encoder

Pada tahap decoding, setiap skenario pengembangan model akan mengimplementasikan algoritma yang berbeda.

```
INPUT:
    - embedding_dim (Integer): Dimensi dari vektor
embedding.
    - units (Integer): Jumlah unit pada lapisan LSTM.
    - vocab_size (Integer): Ukuran kosakata untuk
lapisan embedding.
    - x (Tensor): Token masukan yang direpresentasikan
sebagai indeks dari kosakata.
    - features (Tensor): Fitur gambar yang telah
diekstraksi.
    - hidden (Tensor): hidden state dari model
sebelumnya.

OUTPUT:
    - output (Tensor): Prediksi keluaran setelah
diproses oleh model.
    - state (Tensor): hidden state terbaru setelah LSTM.
    - attention_weights (Tensor): Matriks attention
weight untuk fitur gambar.

1  Class Decoder:
2    Initialize(embedding_dim, units, vocab_size):
3       EmbeddingLayer = Embedding(vocab_size,
                          embedding_dim)
```

```
4       LSTM1 = LSTM(units, activation='tanh',
                  return_sequences=True)
5       LSTM2 = Bidirectional(LSTM(units,
                  activation='tanh',
                  return_sequences=True))
6       LSTM3 = Bidirectional(LSTM(units,
                  activation='tanh',
                  return_sequences=True,
                  return_state=True))
7       Attention = VisualAttention(units)
8       FC1 = Dense(units)
9       FlattenLayer = Flatten()
10      FC2 = Dense(vocab_size)
11
12 Call(x, features, hidden):
13   context_vector, attention_weights = Attention(
                                          features, hidden)
14   x = EmbeddingLayer(x)
15   feat = features[:, :-80, :]
16    x = Concatenate(x, feat)
17    x = LSTM1(x)
18    x = LSTM2(x)
19    x = Slice(x, context_vector.shape[0], :, :-128)
20    x = Concatenate(context_vector, x)
21    output, state, *_ = LSTM3(x)
22    x = FC1(x)
23    x = Reshape(x, (-1, x.shape[2]))
24    x = FC2(x)
25    Return x, state, attention_weights
```

Pseudocode 5.4 Pelatihan Model LSTM

```
INPUT:
    - embedding_dim (Integer): Dimensi dari vektor
embedding.
    - units (Integer): Jumlah unit pada lapisan CNN.
    - vocab_size (Integer): Ukuran kosakata untuk
lapisan embedding.
    - x (Tensor): Token masukan yang direpresentasikan
sebagai indeks dari kosakata.
    - features (Tensor): Fitur gambar yang telah
diekstraksi.
    - hidden (Tensor): hidden state dari model
sebelumnya.

OUTPUT:
    - output (Tensor): Prediksi keluaran setelah
diproses oleh model.
    - state (Tensor): hidden state terbaru setelah CNN.
    - attention_weights (Tensor): Matriks attention
weught untuk fitur gambar.


1  Class Decoder:
2    Initialize(embedding_dim, units, vocab_size):
3       EmbeddingLayer = Embedding(vocab_size,
                          embedding_dim)
4       Conv1 = Conv1D(filters=256, kernel_size=3,
                activation='relu', padding='same')
5       Conv2 = Conv1D(filters=128, kernel_size=3,
                activation='relu', padding='same')
6       GlobalPool = GlobalMaxPooling1D()
7       Attention = VisualAttention(units)
8       FC1 = Dense(units)
```

```
9       FC2 = Dense(vocab_size)
10 Call(x, features, hidden):
11   context_vector, attention_weights = Attention(
                                        features, hidden)
12   x = EmbeddingLayer(x)
13   feat = features[:, :-1, :]
14    feat = Mean(feat, axis=1)
15    x = Concatenate(x, feat)
16    x = Conv1(x)
17    x = Conv2(x)
18    x = GlobalPool(x)
19    x = ExpandDims(x, 1)
20    x = Concatenate(context_vector, x)
21    x = FC1(x)
22    x = Reshape(x, (-1, x.shape[2]))
23    x = FC2(x)
24    Return x, state, attention_weights
```

Pseudocode 5.5 Pelatihan Model CNN

```
INPUT:
    - embedding_dim (Integer): Dimensi dari vektor
embedding.
    - units (Integer): Jumlah unit pada lapisan GRU.
    - vocab_size (Integer): Ukuran kosakata untuk
lapisan embedding.
    - x (Tensor): Token masukan yang direpresentasikan
sebagai indeks dari kosakata.
    - features (Tensor): Fitur gambar yang telah
diekstraksi.
    - hidden (Tensor): hidden state dari model
sebelumnya.
```

```
OUTPUT:
    - output (Tensor): Prediksi keluaran setelah
diproses oleh model.
    - state (Tensor): hidden state terbaru setelah GRU.
    - attention_weights (Tensor): Matriks attention
weight untuk fitur gambar.


1  Class Decoder:
2    Initialize(embedding_dim, units, vocab_size):
3       EmbeddingLayer = Embedding(
                          vocab_size, embedding_dim)
4       GRU1 = GRU(units, activation='tanh',
                    return_sequences=True)
5       GRU2 = Bidirectional(GRU(units,
                activation='tanh',
                return_sequences=True))
6       GRU3 = Bidirectional(GRU(units,
                activation='tanh',
                return_sequences=True,
                return_state=True))
7       Attention = VisualAttention(units)
8       FC1 = Dense(units)
9       FC2 = Dense(vocab_size)
10
11 Call(x, features, hidden):
12   context_vector, attention_weights = Attention(
                                      features, hidden)
13   x = EmbeddingLayer(x)
14   feat = features[:, :-80, :]
15    x = Concatenate(x, feat)
16    x = GRU1(x)
17    x = GRU2(x)
18    x = Slice(x, context_vector.shape[0], :, :-128)
19    x = Concatenate(context_vector, x)
```

```
20    output, state, *_ = GRU3(x)
21    x = FC1(x)
22    x = Reshape(x, (-1, x.shape[2]))
23    x = FC2(x)
24    Return x, state, attention_weights
```

Pseudocode 5.6 Pelatihan Model GRU

Setiap iterasi tahap decoding akan menghasilkan *context vector* untuk iterasi berikutnya melalui serangkaian pengolahan *hidden state* oleh fungsi *attention*. Berikut adalah gambaran algoritma a*ttention* untuk menghasilkan *context vector*.

```
INPUT:
    - units (Integer): Jumlah unit dalam lapisan
perhatian (attention).
    - features (Tensor): Fitur gambar yang telah
diekstraksi.
    - hidden (Tensor): hidden state dari model
sebelumnya.

OUTPUT:
    - context_vector (Tensor): Vektor konteks yang
mewakili bagian penting dari fitur gambar berdasarkan
attention weight.
    - attention_weights (Tensor): Matriks attention
weight yang menunjukkan seberapa relevan setiap bagian
dari fitur gambar.

1   Class VisualAttention:
2     Initialize(units):
3       W1 = Dense(units)
4       W2 = Dense(units)
5       V = Dense(1)
6
```

```
7   Call(features, hidden):
8     hidden_with_time_axis = ExpandDims(hidden, 1)
9     attention_hidden_layer = Tanh(W1(features) +
                              W2(hidden_with_time_axis))
10   score = V(attention_hidden_layer)
11   attention_weights = Softmax(score, axis=1)
12   context_vector = attention_weights * features
13   context_vector = ReduceSum(context_vector, axis=1)
14   Return context_vector, attention_weights
```

Pseudocode 5.7 Pengolahan Context Vector

5.4. Data Testing

Setelah proses pelatihan mesin, tahap selanjutnya adalah pengujian dan evaluasi performa. Setiap skenario akan dievaluasi menggunakan parameter BLEU Score dan ROUGE Score. Berikut adalah algoritma fungsi BLEU Score dan ROUGE Score untuk evaluasi performa model.

```
INPUT:
    - model (Pretrained Model): Model yang telah dilatih
untuk menghasilkan teks.
    - feature (Tensor): Fitur gambar yang digunakan
sebagai input model.
    - references (List of Strings): Daftar kalimat
referensi untuk evaluasi BLEU.
    - vocab_size (Integer): Ukuran kosakata yang
digunakan dalam model.

OUTPUT:
    - BLEU-1, BLEU-2, BLEU-3, BLEU-4 (Floats): Skor BLEU
untuk mengukur kualitas teks yang dihasilkan.
```

```
1   candidate = ‘startseq’
2   FOR i in vocab_size = 32 :
3     Prediction = model.predict (feature)
4     candidate.append(Prediciton)
5     IF prediction == ‘endseq’ :
6        break
7   delete.cadidate([‘startseq’, ‘endseq’])
8   BLEU-1 = BLEU(references, candidate,
                  weights(1, 0, 0, 0))
9   BLEU-2 = BLEU(references, candidate,
                  weights(0.5, 0.5, 0, 0))
10  BLEU-3 = BLEU(references, candidate,
                  weights(0.33, 0.33, 0.33, 0))
11  BLEU-4 = BLEU(references, candidate,
                  weights(0.25, 0.25, 0.25, 0.25))
12 Calculate mean_of_BLEU
13 Tuning model
14 EXPORT evaluated model
```

Pseudocode 5.8 Pengujian dan Evaluasi Performa BLEU Score

```
INPUT:
    - candidates (List of Strings): Daftar teks yang
dihasilkan oleh model untuk dievaluasi.
    - references (List of Strings): Daftar kalimat
referensi yang digunakan untuk evaluasi.

OUTPUT:
    - Rouge-1, Rouge-2, Rouge-L, Rouge-Lsum (Floats):
Skor evaluasi Rouge untuk mengukur kesamaan antara hasil
dan referensi.

1  def rouge_score():
2     rouge = evaluate.load('rouge')
3     result = rouge.compute(predictions=candidates,
             references=references)
```

```
4     print(f'Rouge-1 = {result["rouge1"]*100}%\n')
5     print(f'Rouge-2 = {result["rouge2"]*100}%\n')
6     print(f'Rouge-L = {result["rougeL"]*100}%\n')
7     print(f'Rouge-Lsum ={result["rougeLsum"]*100}%\n')
```

Pseudocode 5.9 Pengujian dan Evaluasi Performa ROUGE Score

[Halaman ini sengaja dikosongkan]

BAB VI
PENGUJIAN DAN EVALUASI

Bab ini menjelaskan tahap uji coba model yang telah dikembangkan untuk memastikan fungsionalitasnya. Pengujian dilakukan guna memverifikasi kesesuaian hasil implementasi dengan analisis dan desain yang telah dibuat. Fokus pengujian mencakup kesesuaian deskripsi oleh model terhadap deskripsi yang telah dibuat oleh sumber lainnya. Setiap hasil pengujian dibandingkan dengan memanfaatkan BLEU Score dan ROUGE Score. Langkah ini memastikan hasil deskripsi oleh model sesuai kebutuhan dan mencapai tujuan pengembangan.

6.1. Tujuan Pengujian

Pengujian dilakukan terhadap model utntuk menguji kemampuan model dalam mendeskripsikan sebuah gambar. Hasil pengujian setiap skenario akan dibandingkan dengan skenario lainnya untuk mengetahui metode *deep learning* terbaik dalam pengembangan model *image captioning*. Selain itu, pada skenario 1 terdapat pengujian menggunakan dataset lama untuk mengetahui pengaruh metode pelabelan dataset baru terhadap metode pelabelan dataset lama.

6.2. Kriteria Pengujian

Penilaian atas pencapaian tujuan pengujian didapatkan dengan membandingkan hasil deskripsi oleh model terhadap deskripsi dari sumber lainnya yang diberikan pada tahap pelabelan dataset. Pengujian dan evaluasi performa memanfaatkan BLEU score dari BLEU-1 hingga BLEU-4 dengan menyesuaikan bobot kalimat dari 1-gram

hingga 4-gram. Serta hasil ROUGE-n dan ROUGE-L, dimana ROUGE-n yang digunakan adalah ROUGE-1 dan ROUGE-2 yang merepresentasikan 1-gram dan 2-gram, sedangkan ROUGE-L yang digunakan adalah ROUGE-L dan ROUGE-Lsum. Sebuah model akan dikatakan baik jika rata-rata BLEU score dan ROUGE score dari keseluruhan data uji memiliki nilai tinggi.

6.3. Skenario Pengujian

Skenario pengujian setiap model dilakukan dengan membandingkan hasil deskripsi setiap citra oleh model terhadap deskripsi dari sumber lainnya. Langkah-langkah pengujian yaitu sebagai berikut :

1. Setiap gambar yang digunakan sebagai bahan uji akan dibuatkan deskripsi oleh model.
2. Setiap deskripsi oleh model akan dievaluasi kesesuaiannya terhadap semua kalimat deskripsi pada tahap pelabelan dataset.
3. Menghitung BLEU score dari BLEU-1 hingga BLEU-4 dengan menyesuaikan bobot kalimat dari 1-gram hingga 4-gram.
4. Dari semua BLEU score yang dihasilkan, akan dihitung rata-rata untuk setiap tingkat BLEU score.
5. Rata-rata BLEU score akan mewakili keseluruhan kualitas model yang telah dibuat.
6. Hal yang sama juga diterapkan pada ROUGE score dengan menggunakan ROUGE-n dengan jenis ROUGE-1 dan ROUGE-2, ROUGE-L, serta ROUGE-Lsum

Terdapat 3 jenis evaluasi yang dijalankan, yaitu berdasarkan metode pelabelan dataset, metode pemodelan, serta hyperparameter tuning. Evaluasi pertama merupakan

evauasi performa untuk membandingkan metode pelabelan dataset. Pada evaluasi ini, skenario pemodelan yang digunakan adalah skenario 1 atau skenario LSTM. Terdapat 3 dataset yang diujikan dengan spesifikasi pada Tabel 6.1. Ketiga dataset akan digunakan untuk membuat model *image captioning* dan hasil pengujiannya akan dibandingkan untuk menentukan metode pelabelan terbaik berdasarkan BLEU Score dan ROUGE.

Tabel 6. 1 Perbandingan Dataset

Pembanding	**Dataset 1**	**Dataset 2**	**Dataset 3**
Jumlah citra	3140	1447	3001
Jumlah caption untuk setiap citra	5	6	4 - 6
Deskripsi citra	Terdiri dari 1693 gambar dari dataset flickr30k serta 1447 gambar trotoar	Terdiri dari gambar trotoar yang sama pada Dataset 1	Terdiri dari 1447 gambar trotoar yang sama pada Dataset 1 serta 1554 gambar sekitar kampus ITS surabaya

Metode pelabelan	Dilakukan oleh 1 orang saja	Label tiap gambar terdiri dari 3 label baru yang dibuat oleh 3 orang yang berbeda, 2 label dari Dataset 1, serta 1 label dari GPT	Label gambar trotoar merupakan 6 label yang sama pada Dataset 2, sedangkan gambar sekitar kampus ITS terdiri dari 3 label yang dibuat oleh 3 orang yang berbeda serta 1 label dari GPT

Evaluasi kedua merupakan evaluasi performa untuk menentukan metode pemodelan terbaik. Pada evaluasi ini akan dilakukan uji performa skenario 1, 2, dan 3. Dataset yang digunakan pada evaluasi ini adalah Dataset 3, terlepas dari hasil dari evaluasi pertama. Fungsi aktivasi yang digunakan pada decoder untuk setiap skenario pada evaluasi ini adalah tanh untuk skenario 1 dan 2, serta relu untuk skenario 3. Untuk optimizer, digunakan optimizer Adam untuk semua skenario.

Evaluasi ketiga adalah evaluasi hyperparameter tuning pada semua skenario. Pada evaluasi ini, parameter

yang akan diujikan adalah fungsi aktifasi pada bagian decoder. Fungsi aktifasi yang diujikan adalah tanh, relu, dan sigmoid. Dataset yang digunakan pada evaluasi ini adalah Dataset 3, sama seperti evaluasi kedua. Optimizer yang digunakan pada decoder untuk semua skenario pada evaluasi ini berupa optimizer Adam.

6.4. Evaluasi Pengujian

Hasil evaluasi pertama, evaluasi performa model untuk menentukan metode pelabelan dataset terbaik, ditunjukkan pada Tabel 6.2 dan Tabel 6.3.

Tabel 6. 2 Perbandingan BLEU Score untuk Evaluasi Metode Pelabelan Dataset

Tingkat BLEU Score	Dataset 1	Dataset 2	Dataset 3
BLEU-1	51.49%	40.84%	23.12%
BLEU-2	32.14%	23.91%	8.65%
BLEU-3	15.85%	12.29%	2.74%
BLEU-4	7.29%	5.19%	0.87%

Tabel 6. 3 Perbandingan Nilai ROUGE untuk Evaluasi Metode Pelabelan Dataset

ROUGE	Dataset 1	Dataset 2	Dataset 3
ROUGE-1	40.05%	34.56%	21.75%
ROUGE-2	18.66%	14.33%	5.00%
ROUGE-L	35.25%	30.96%	18.51%
ROUGE-Lsum	35.23%	30.95%	18.49%

Hasil pengujian kedua, evaluasi performa model berdasarkan metode *deep learning* yang diimplementasikan, dapat dilihat pada Tabel 6.4 dan Tabel 6.5.

Tabel 6. 4 Perbandingan BLEU Score untuk Evaluasi Metode Pemodelan

Tingkat BLEU Score	**LSTM**	**CNN**	**GRU**
BLEU-1	23.12%	25.56%	23.05%
BLEU-2	8.65%	10.90%	8.30%
BLEU-3	2.74%	2.98%	2.00%
BLEU-4	0.87%	0.65%	0.47%

Tabel 6. 5 Perbandingan Nilai ROUGE untuk Evaluasi Metode Pemodelan

ROUGE	**LSTM**	**CNN**	**GRU**
ROUGE-1	21.75%	23.44%	21.63%
ROUGE-2	5.00%	5.99%	4.86%
ROUGE-L	18.51%	19.96%	18.59%
ROUGE-Lsum	18.49%	19.94%	18.58%

Hasil evaluasi ketiga, hyperparameter tuning fungsi aktifasi pada decoder semua skenario, dapat dilihat pada Tabel 6.6 dan Tabel 6.7.

Tabel 6. 6 Perbandingan BLEU Score untuk Pengujian Fungsi Aktifasi

		BLUE-1	**BLUE-2**	**BLUE-3**	**BLUE-4**
LSTM	tanh	23.12%	8.65%	2.74%	0.87%
	relu	23.09%	8.10%	2.24%	0.43%

	sigmoid	21.51%	7.63%	1.70%	0.29%
GRU	tanh	23.05%	8.30%	2.00%	0.47%
	relu	22.18%	8.83%	2.49%	0.68%
	sigmoid	20.88%	7.45%	1.78%	0.28%
CNN	tanh	21.41%	8.40%	2.31%	0.24%
	relu	25.56%	10.90%	2.98%	0.65%
	sigmoid	21.50%	7.37%	1.69%	0.33%

Tabel 6. 7 Perbandingan ROUGE Score untuk Pengujian Fungsi Aktifasi

		ROUGE-1	**ROUGE-2**	**ROUGE-L**	**ROUGE-Lsum**
LSTM	tanh	21.75%	5.00%	18.51%	18.49%
	relu	21.87%	4.64%	19.14%	19.19%
	sigmoid	20.49%	4.32%	17.40%	17.41%
GRU	tanh	21.63%	4.86%	18.59%	18.58%
	relu	21.80%	5.37%	18.77%	18.74%
	sigmoid	20.62%	4.32%	17.67%	17.68%
CNN	tanh	21.10%	4.95%	18.25%	18.24%
	Relu	23.44%	5.99%	19.96%	19.94%
	sigmoid	20.09%	4.15%	17.29%	17.26%

6.5. Pembahasan

Berdasarkan hasil perhitungan BLEU score dan ROUGE score, evaluasi pertama dilakukan terhadap metode pelabelan dataset. Dataset yang menghasilkan performa model terbaik adalah Dataset 1, dengan score BLUE-1 51.49%, BLUE-2 32.14%, BLUE-3 15.85%, BLUE-4 7.29%, ROUGE-1 40.05%, ROUGE-2 18.66%, ROUGE-L 35.25%, ROUGE-Lsum 35.23%, dalam pengujian menggunakan metode LSTM dan fungsi aktifasi Adam. Kualitas dataset image captioning sangat dipengaruhi oleh beberapa faktor, seperti variasi objek dalam gambar, detail dan variasi label, serta jumlah gambar yang tersedia. Dataset 2 memiliki jumlah gambar paling sedikit, dengan 1447 total gambar, yang berkontribusi terhadap rendahnya performa model. Sementara itu, meskipun Dataset 3 memiliki jumlah gambar yang hampir sama dengan Dataset 1, kualitas label yang kurang detail dan bervariasi menyebabkan performa model menjadi lebih buruk. Dengan demikian, dapat disimpulkan bahwa keberagaman label serta jumlah gambar yang memadai menjadi aspek penting dalam meningkatkan performa model image captioning.

Evaluasi kedua dilakukan terhadap metode pemodelan yang digunakan. Dari hasil penelitian, metode CNN menghasilkan model dengan performa terbaik, dengan BLEU-1 sebesar 25.56%, BLEU-2 sebesar 10.90%, BLEU-3 sebesar 2.98%, dan BLEU-4 sebesar 0.65%. Sementara itu, hasil ROUGE score menunjukkan bahwa ROUGE-1 memiliki nilai 23.44%, ROUGE-2 sebesar 5.99%, ROUGE-

L sebesar 19.96%, dan ROUGE-Lsum sebesar 19.94%. Pengujian ini dilakukan menggunakan Dataset 3 dengan fungsi aktivasi Adam. Selain itu, terdapat indikasi bahwa semakin panjang atau semakin kompleks label suatu gambar, maka performa model cenderung menurun. Hal ini disebabkan oleh metode perhitungan BLEU score dan ROUGE score yang membandingkan kesamaan kata antara caption yang dihasilkan model dengan label pada dataset. Oleh karena itu, diperlukan keseimbangan dalam menentukan panjang label untuk menghindari distorsi dalam perhitungan evaluasi.

Evaluasi ketiga berkaitan dengan hyperparameter tuning, yang dalam penelitian ini difokuskan pada fungsi aktivasi. Fungsi aktivasi yang memberikan performa terbaik adalah Relu, dengan rata-rata BLEU-1 sebesar 23.61%, BLEU-2 sebesar 9.28%, BLEU-3 sebesar 2.57%, dan BLEU-4 sebesar 0.59%. Sementara itu, hasil ROUGE score menunjukkan bahwa ROUGE-1 memiliki nilai 22.37%, ROUGE-2 sebesar 5.33%, ROUGE-L sebesar 19.29%, dan ROUGE-Lsum sebesar 19.29%. Namun, penelitian ini hanya menguji fungsi aktivasi sebagai satu-satunya hyperparameter, sedangkan setiap metode atau algoritma memiliki berbagai hyperparameter lain yang juga dapat mempengaruhi performa model. Oleh karena itu, hasil evaluasi hyperparameter dalam penelitian ini masih bersifat terbatas dan memerlukan eksplorasi lebih lanjut untuk mendapatkan konfigurasi optimal bagi setiap metode pemodelan yang digunakan.

[Halaman ini sengaja dikosongkan]

BAB VII
KESIMPULAN DAN SARAN

7.1. Kesimpulan

Berdasarkan hasil evaluasi yang telah dilakukan, dapat disimpulkan bahwa kualitas dataset, metode pengembanga model yang digunakan, serta pemilihan fungsi aktivasi sangat berpengaruh terhadap performa *model image captioning*. Dalam perbandingan metode pelabelan dataset, metode yang diterapkan pada Dataset 1 terbukti menghasilkan performa terbaik jika dibandingkan dengan metode pelabelan Dataset 2 dan Dataset 3. Sementara itu, metode *deep learning* CNN menunjukkan performa terbaik jika dibandingkan dengan model LSTM dan GRU. Penggunaan fungsi aktivasi Relu juga menunjukkan hasil yang lebih unggul dibandingkan alternatif lainnya.

7.2. Saran

Terdapat beberapa faktor yang masih perlu diperhatikan, seperti jumlah dan kualitas label dalam dataset, variasi hasil caption yang dihasilkan model, serta keterbatasan dalam pengujian *hyperparameter*. Selain itu, metode evaluasi yang digunakan masih memiliki kelemahan karena skenario pengujian hanya dilakukan sekali tanpa perulangan yang cukup. Oleh karena itu, penelitian di masa depan diharapkan mampu mempertimbangkan optimasi *hyperparameter* yang lebih luas, penerapan batasan label, serta pendekatan evaluasi yang lebih akurat untuk meningkatkan performa model.

[Halaman ini sengaja dikosongkan]

DAFTAR PUSTAKA

Castro, R., Pineda, I., Lim, W., & Morocho-Cayamcela, M. E. (2022). Deep Learning Approaches Based on Transformer Architectures for Image Captioning Tasks. *IEEE Access*, *10*, 33679–33694. https://doi.org/10.1109/ACCESS.2022.3161428

de Rautlin de la Roy, E., Recht, T., Zemmari, A., Bourreau, P., & Mora, L. (2023). Deep learning models for building window-openings detection in heating season. *Building and Environment*, *231*, 110019. https://doi.org/10.1016/j.buildenv.2023.110019

Ghandi, T., Pourreza, H., & Mahyar, H. (2022). *Deep Learning Approaches on Image Captioning: A Review*. https://doi.org/10.1145/3617592

Kesiman, Antara, M. W., & Dermawan, K. T. (2021). AKSALont: Automatic Transliteration Application for Balinese Palm Leaf Manuscripts with LSTM Model. *Jurnal Teknologi Dan Sistem Komputer 9*, *3*, 142–149.

Kumari, L., Debarma, S., & Kar, N. (2011). *Machine translation evaluation system: development & comparison. National Institute of Technology*.

Lavie, A. (2011). Evaluating the Output of Machine Translation Systems. *Association for Machine Translation in the Americas*.

Li, C., Li, G., Wang, K., & Han, B. (2022). A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient

integrated energy systems. *Energy*, *259*, 124967. https://doi.org/10.1016/j.energy.2022.124967

Li, R., Liang, H., Shi, Y., Feng, F., & Wang, X. (2020). Dual-CNN: A Convolutional language decoder for paragraph image captioning. *Neurocomputing*, *396*, 92–101. https://doi.org/10.1016/j.neucom.2020.02.041

Navastara, D. A., Ansori, D. B., Suciati, N., & Akbar, Z. F. (2023). Combination of DenseNet and BiLSTM Model for Indonesian Image Captioning. *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, 994–999. https://doi.org/10.1109/ICAMIMIA60881.2023.10427729

Sanchez-Gomez, J. M., Vega-Rodríguez, M. A., & Pérez, C. J. (2022). A multi-objective memetic algorithm for query-oriented text summarization: Medicine texts as a case study. *Expert Systems with Applications*, *198*, 116769. https://doi.org/10.1016/j.eswa.2022.116769

Zhang, J., Li, K., & Wang, Z. (2021). Parallel-fusion LSTM with synchronous semantic and visual information for image captioning. *Journal of Visual Communication and Image Representation*, *75*, 103044. https://doi.org/10.1016/j.jvcir.2021.103044

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, *42*, 146–157. https://doi.org/10.1016/j.inffus.2017.10.006

BIODATA PENULIS I

Nama	: Wardatul Amalia Safitri
Tempat, Tanggal Lahir	: Gresik, 15 April 2003
Jenis Kelamin	: Perempuan
Telepon	: 08979610029
Email	: 5025211006@student.its.ac.id

AKADEMIS

Kuliah	: Departemen Teknik Informatika – FTEIC , ITS
Angkatan	: 2021
Semester	: 7 (Tujuh)

BIODATA PENULIS II

Nama	: Hammuda Arsyad
Tempat, Tanggal Lahir	: Tegal, 17 Januari 2003
Jenis Kelamin	: Laki laki
Telepon	: 08113597728
Email	: 5025211146@student.its.ac.id

AKADEMIS

Kuliah	: Departemen Teknik Informatika – FTEIC , ITS
Angkatan	: 2021
Semester	: 7 (Tujuh)