KAPSYEN AUTOMATIK IMEJ DALAM DWIBAHASA

NUR SAHIRA SOFEA BINTI AZIZAN

TS. DR. WAN FARIZA BINTI PAIZI @ FAUZI

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

ABSTRAK

Projek inovatif ini mengadopsi pendekatan dua fasa untuk meningkatkan adaptabiliti model kapsyen imej. Pada mulanya, kapsyen dihasilkan dalam Bahasa Inggeris menggunakan model kapsyen imej standard dengan VGG-16, dan kemudian diterjemahkan dengan lancar ke dalam Bahasa Melayu menggunakan API Google Translate. Projek ini melibatkan lima fasa utama: Permulaan Projek, Pengumpulan Data, Pembangunan Algoritma, Pengujian Model, dan Pelaksanaan Model.Dengan menterjemahkan kapsyen ke dalam Bahasa Melayu, projek ini memastikan inklusiviti dan aksesibiliti, meningkatkan pengalaman pengguna dalam pelbagai konteks linguistik.Model ini dinilai menggunakan metrik Bilingual Evaluation Understudy (BLEU), mencapai skor BLEU-1 sebanyak 0.540513 dan skor BLEU-2 sebanyak 0.323107 pada epoch terbaiknya (epoch 25), dengan skor purata BLEU-1 sebanyak 0.514348 dan skor purata BLEU-2 sebanyak 0.291748 merentasi semua epoch. Maklum balas pengguna menunjukkan 61.8% terjemahan dinilai sebagai "Sangat Tepat". Walaupun terdapat beberapa cabaran seperti ketersediaan set data dalam Bahasa Melayu yang terhad dan kekangan memori, projek ini telah memberikan sumbangan yang signifikan kepada bidang kapsyen imej automatik. Penambahbaikan masa depan termasuk memperluaskan set data, mengoptimumkan kualiti kapsyen, dan meningkatkan interaktiviti antara muka pengguna.

Kata kunci: kapsyen imej, pembelajaran mendalam, terjemahan automatik, Bahasa Melayu, BLEU, VGG-16, Google Translate.

PENGENALAN

Dalam era media digital, projek "Kapsyen Automatik Imej dalam Dwibahasa" adalah sebuah projek yang menggabungkan teknologi penglihatan komputer *(Computer Vision)* dan pemprosesan bahasa semula jadi *(NLP).* Elemen penting dalam projek ini ialah kapsyen imej, yang melibatkan penjanaan penerangan teks untuk imej. Penerangan imej yang dihasilkan akan tersedia dalam bahasa Inggeris atau bahasa Melayu ataupun kedua-duanya.

Pada asasnya, sistem ini menggunakan algoritma dan teknik pembelajaran mendalam

untuk menganalisis isi imej yang kompleks. Bidang penglihatan komputer, yang merupakan sebahagian daripada kecerdasan buatan yang lebih meluas membolehkan sistem "melihat" dan memahami apa yang ada dalam imej. Ia boleh mengenal pasti objek, pemandangan, dan elemen visual yang lain. Selain itu, pemprosesan bahasa semula jadi dapat membantu sistem "bercakap" dalam bahasa manusia dengan menukarkan tafsiran visual kepada teks yang logik dan bermakna.

Penciptaan kapsyen imej automatik telah menjadi alat penting dalam memudahkan akses dan pengurusan kandungan digital, membantu dalam pelbagai tugas dari pengambilan kandungan hingga membantu pengguna yang mempunyai masalah penglihatan. Namun, pembangunan sistem penciptaan kapsyen imej yang kuat yang dapat mengendalikan ouput dwibahasa terutamanya untuk Bahasa Melayu dan Inggeris yang menghadapi cabaran yang signifikan. Ini termasuk keperluan untuk set data yang luas dan pelbagai, kompleksiti dalam terjemahan yang tepat antara bahasa dengan struktur sintaks yang berbeza, serta integrasi kapsyen yang relevan dengan konteks yang konsisten secara budaya dan linguistik.

Objektif projek ini adalah untuk merancang dan membangun antara muka pengguna grafik (GUI) bagi penjanaan kapsyen imej dwibahasa secara automatik dengan menggunakan senibina rangkaian neural Visual Geometry Group 16 (VGG-16). Projek ini melibatkan pembangunan model menggunakan dataset Flickr8k yang menekankan aktiviti manusia. Fokus utama adalah pada pelatihan dan penilaian model bagi memastikan kemampuan menghasilkan kapsyen yang deskriptif dan relevan.

Selain itu, ketepatan sistem penjanaan kapsyen akan dinilai menggunakan metrik Bilingual Evaluation Understudy (BLEU). Akhir sekali, projek ini akan melakukan pengujian ketepatan terjemahan untuk memastikan kapsyen dalam Bahasa Inggeris yang diterjemahkan ke Bahasa Melayu adalah sesuai dari segi konteks dan bahasa.

METODOLOGI KAJIAN

Model pembangunan mesin atau "Machine Learning Model" adalah aspek penting dalam menyelesaikan masalah yang berkaitan dengan imej. Proses ini melibatkan beberapa fasa utama iaitu Fasa Penyediaan, Fasa Pembangunan Algorima, Fasa Pengujian serta Fasa Implimentasi Model yang kesemuanya kritikal untuk mencapai hasil yang diinginkan dalam projek ini.

Fasa Penyediaan Data

Fasa pertama projek ini adalah pengumpulan data. Data yang diperlukan, termasuk imej dan maklumat berkaitan. Data yang digunakan iaitu Flickr8k[1] digunakan untuk melatih model pembelajaran mesin dalam projek ini. Proses ini melibatkan penyediaan data seperti membersihkan data, serta menyelaraskan format data supaya sesuai dengan penggunaan dalam model pembelajaran mesin. Data yang bersih dan teratur adalah asas untuk

[1] https://www.kaggle.com/datasets/adityajn105/flickr8k

menghasilkan model yang tepat dan boleh diandalkan.

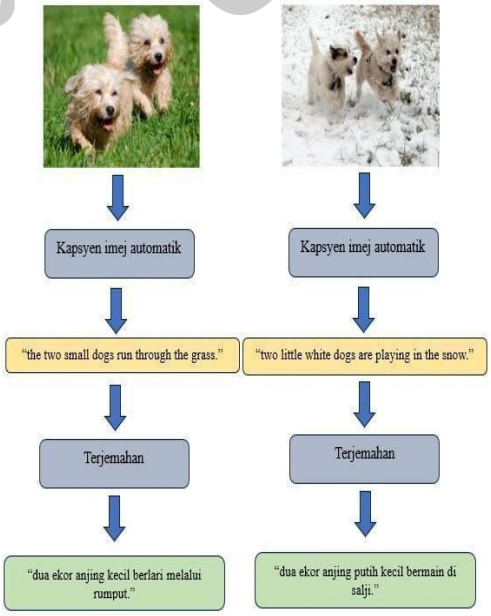
Fasa Pembangunan Algoritma

Fasa Pembangunan Algoritma adalah satu peringkat penting di mana asas untuk pengekapsyen imej automatik dan terjemahan diletakkan. Pada permulaan, data input mentah, biasanya imej, mengalami siri langkah pemprosesan, memastikan keseragaman, standardisasi, dan relevansi untuk model pembelajaran mesin. Input yang telah diproses ini berfungsi sebagai titik permulaan untuk peringkat berikutnya.

Perjalanan bermula dengan pengekapsyen imej automatik, di mana model mengekstrak ciri-ciri bermakna dari imej input dan menghasilkan kapsyen deskriptif dalam Bahasa Inggeris. Fasa ini menumpukan kepada pemanfaatan rangkaian neural canggih, seperti rangkaian neural konvolusional (CNNs) dan rangkaian neural ulang (RNNs), untuk memahami dan menyuarakan kandungan imej.

Selepas penghasilan kapsyen Bahasa Inggeris, proses ini dengan lancar berpindah ke fasa terjemahan. Kapsyen Bahasa Inggeris dimasukkan ke dalam API terjemahan, memudahkan transformasi teks ke dalam Bahasa Melayu. Integrasi mekanisme terjemahan ini membolehkan penciptaan kapsyen dalam pelbagai bahasa, meningkatkan aksesibiliti dan kegunaan.

Pada peringkat ini sistem telah menerima imej sebagai input dan menghasilkan kapsyen secara automatik dalam Bahasa Inggeris. Seterusnya, untuk menghasilkan kapsyen dalam Bahasa Melayu, proses penterjemahan akan dilakukan kepada kapsyen di dalam model tersebut mengunakan API terjemahan. Reka Bentuk Algoritma boleh dilihat seperti di Rajah 1.



Rajah 1 Reka Bentuk Algoritma

i. Kapsyen Imej Automatik

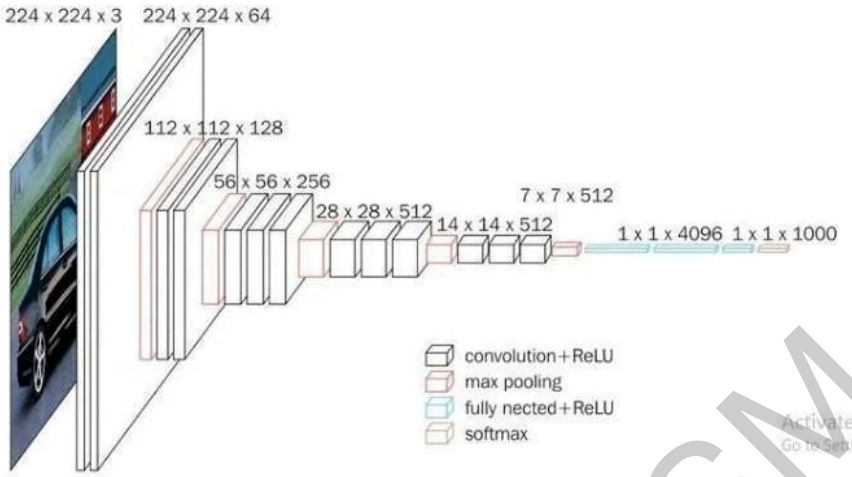
Tugas pemberian kapsyen imej boleh dibahagikan kepada dua model: Model Berdasarkan Imej mengekstrak ciri-ciri dari imej, sementara Model Berdasarkan Bahasa mencipta ayat tentang imej menggunakan kapsyen sebelumnya bersama dengan ciri-ciri yang diberikan oleh model berdasarkan imej.

Aspek pertama yang penting adalah pemilihan senibina rangkaian neural yang sesuai. Model seperti gabungan *Convolutional Neural Networks (CNNs)* untuk pengeluaran ciri imej dan *Recurrent Neural Networks (RNNs*) atau Transformers untuk penghasilan kapsyen sering digunakan dalam tugas kapsyen imej.

A) Model Berdasarkan Imej - Rangkaian Neural Konvolusi (CNN)

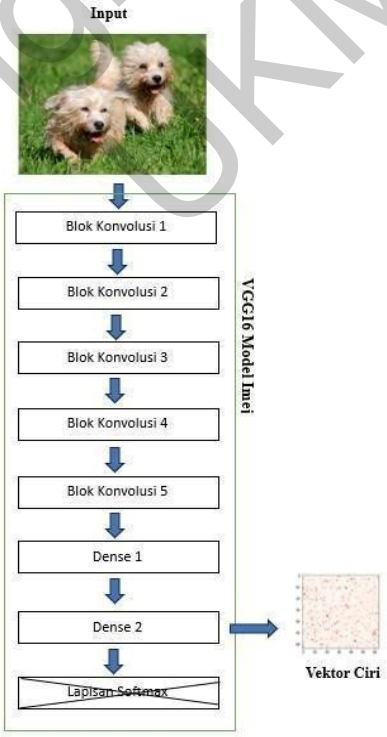
Dalam projek ini, VGG16 (*Visual Geometry Group 16*) akan digunakan sebagai model berdasarkan imej. VGG16 ialah algoritma pengesanan dan klasifikasi objek yang mampu mengklasifikasikan 1000 imej daripada 1000 kategori yang berbeza dengan ketepatan sebanyak 92.7%. Ia merupakan salah satu algoritma popular untuk klasifikasi imej dan mudah digunakan dengan pembelajaran pemindahan.

VGG16 dinamakan demikian kerana ia mempunyai 16 lapisan dengan pemberat, terdiri daripada tiga belas lapisan konvolusi, lima lapisan *max pooling*, dan tiga lapisan *dense*, menjadikan jumlah keseluruhan 21 lapisan, namun ia hanya mempunyai enam belas lapisan pemberat, setiap satu dengan parameter yang boleh dipelajari. Secara ketara, ia menerima input tensor berukuran 224x224 dengan 3 saluran RGB, dan ciri uniknya terletak pada keutamaan lapisan konvolusi filter 3x3 dengan langkah 1, sentiasa menggunakan pembungkus yang sama, dan menggunakan *lapisan max pool filter* 2x2 dengan langkah 2, tersusun secara konsisten sepanjang senibina, termasuk Conv-1 dengan 64 penapis, Conv-2 dengan 128 penapis, Conv-3 dengan 256 penapis, dan Conv-4 serta Conv-5 dengan 512 penapis, diikuti oleh tiga lapisan sepenuhnya tersambung, dengan dua yang pertama mempunyai 4096 saluran masing-masing, dan yang ketiga melakukan klasifikasi ILSVRC 1000-ara dengan 1000 saluran, diakhiri dengan lapisan softmax. Arkitektur VGG16 boleh dilihat mengikut Rajah 2.



Rajah 2 Arkitektur VGG-16 (sumber: https://acesse.dev/dNeFw)

Untuk tujuan penjanaan kapsyen imej, lapisan klasifikasi akhir model VGG-16 dibuang, meninggalkan hanya lapisan konvolusi. Lapisan-lapisan ini, yang tersusun secara hierarki, berperanan sebagai penyaring ciri yang kuat. Lapisan-lapisan bawah menangkap ciri-ciri asas seperti bucu dan tekstur, sementara lapisan-lapisan yang lebih tinggi mengimejkan corak yang lebih kompleks dan representasi objek. Langkah ekstraksi ciri ini menekankan kepentingan mengekalkan maklumat visual yang bernilai yang terkandung dalam lapisan-lapisan konvolusi.



Rajah 3 VGG-16 Model Imej

B) Model Berdasarkan Bahasa - Rangkaian Neural Berulang (RNN)

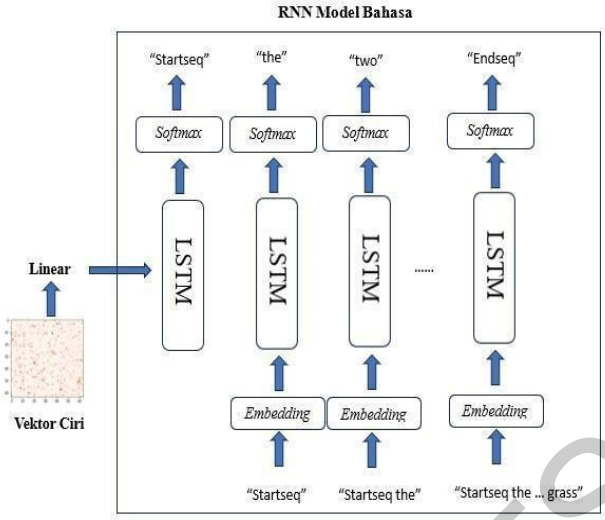
Ciri visual yang diekstrak kemudiannya diintegrasikan ke dalam model berdasarkan bahasa, iaitu rangkaian neural berulang (RNN). Ini berfungsi sebagai jambatan antara domain visual dan tekstual. Model bahasa mengambil ciri visual ini dan menghasilkan kapsyen deskriptif untuk imej, dengan berkesan mempelajari untuk mengaitkan corak visual yang dipelajari dengan huraian tekstual yang sepadan.

Proses menghasilkan jujukan kapsyen melibatkan penggunaan lapisan *embedding* perkataan diikuti oleh lapisan *Long Short-Term Memory* (LSTM). Lapisan embedding perkataan memainkan peranan penting dalam menukar teks input kepada representasi nombor yang dikenali sebagai embedding perkataan. Lapisan ini memetakan perkataan kepada vektor berdimensi tinggi, menangkap hubungan semantik dan maklumat kontekstual. Penggunaan ruang vektor berterusan meningkatkan keupayaan model untuk beroperasi dengan representasi yang bermakna dan kaya konteks.

Berkenaan lapisan LSTM, ia mempunyai peranan khusus sebagai jenis rangkaian neural berulang (RNN) yang direka khas untuk mengatasi cabaran berkaitan dengan pembelajaran ketergantungan jangka panjang dalam data berurutan. LSTM menangani masalah seperti kehilangan dan letupan gradien dengan memasukkan set berat dalaman yang lebih rumit. Kompleksiti ini membolehkan LSTM untuk lebih baik menangkap dan mengingati ketergantungan jangka panjang dalam jujukan, menjadikannya kurang rentan terhadap isu seperti kehilangan dan letupan gradien.

Dalam lapisan LSTM, pengintegrasian sel ingatan adalah ketara. Sel-sel ini memudahkan penyimpanan dan pengambilan maklumat dalam jujukan yang panjang, membolehkan model mengekalkan konteks dan menangkap ketergantungan di sepanjang langkah masa yang berbeza. Mekanisme ingatan ini penting untuk memastikan penyelamatan konteks yang relevan sepanjang proses penjanaan kapsyen, terutamanya dalam tugas di mana urutan perkataan memegang makna yang signifikan, seperti dalam penjanaan kapsyen.

RNN tradisional sering menghadapi kesukaran dalam mempelajari ketergantungan melintang jujukan yang panjang disebabkan masalah kehilangan gradien. LSTM, dengan senibina khasnya, cemerlang dalam pembelajaran ketergantungan jangka panjang dengan meningkatkan keupayaan model untuk menangkap hubungan yang halus di sepanjang jujukan yang panjang. Selain itu, LSTM direka khusus untuk kurang peka terhadap panjang jujukan input, menjadikannya sesuai untuk tugas di mana panjang jujukan boleh berubah-ubah, seperti dalam penjanaan kapsyen untuk imej dengan kompleksiti yang berbeza. Secara ringkasnya, gabungan lapisan embedding perkataan dan LSTM, dengan ciri-ciri khas mereka, membentuk kerangka yang kukuh untuk penjanaan berkesan jujukan kapsyen dalam tugas yang memerlukan pemahaman halus terhadap data berurutan.



Rajah 4 RNN Model Bahasa

C) Pengekod – Penyahkod

Model pengkod-penyahkod digunakan dalam tugas penjanaan kapsyen imej, dengan menggunakan gabungan Rangkaian Neural Konvolusi (CNN) dan Rangkaian Neural Berulang (RNN), khususnya Long Short-Term Memory (LSTM). Dalam konfigurasi ini, rangkaian VGG-16 berfungsi sebagai pengkod. Ia memproses imej masukan untuk mengekstrak ciri-ciri visual yang mengandungi butiran penting yang diperlukan untuk penjanaan kapsyen. Proses ekstraksi ciri ini adalah peranan utama pengkod, yang mana ia mengkompres data masukan yang kompleks (imej) menjadi representasi yang ringkas dan bermakna (vektor ciri).

Selepas proses pengkodan, rangkaian LSTM berfungsi sebagai penyahkod. Ia mengambil ciri-ciri visual yang disediakan oleh pengkod VGG-16 dan menterjemahkannya menjadi deskripsi teks, menghasilkan kapsyen imej secara berurutan. Transformasi dari bentuk data yang dikompres kembali ke output yang berguna (teks) menunjukkan fungsi utama penyahkod dalam model pengkod-penyahkod. Kelebihan menggunakan LSTM dalam tugas ini sangat signifikan kerana ia mahir dalam mengendalikan dependensi jangka panjang, yang penting untuk menghasilkan kapsyen yang koheren dan akurat secara kontekstual.

Integrasi pengkod (VGG-16) dan penyahkod (LSTM) dalam model ini adalah contoh klasik dari arkitektur pengkod-penyahkod, yang sangat penting dalam tugas yang memerlukan terjemahan satu bentuk informasi ke bentuk lain, seperti menterjemahkan data visual ke dalam teks dalam penjanaan kapsyen imej. Penyusunan model ini tidak hanya menunjukkan interaksi antara komponen pengkod dan penyahkod tetapi juga menonjolkan peranan elemen pemodelan bahasa seperti penyematan kata dan lapisan LSTM dalam memproses dan menghasilkan data berurutan, memastikan output akhir bermakna dan relevan dengan imej masukan.

ii. Terjemahan

Selepas menghasilkan kapsyen Bahasa Inggeris, model berinteraksi dengan API terjemahan untuk mendapatkan versi terjemahan dalam bahasa Melayu. Langkah tambahan ini membolehkan model menjadi fleksibel dan boleh menyesuaikan diri dengan pelbagai bahasa tanpa perlu dilatih secara eksplisit dalam setiap bahasa.

Untuk projek ini, API terjemahan yang telah dipilah ialah API Penterjemahan Google Cloud API Penterjemahan Google Cloud menawarkan beberapa kelebihan yang unik yang menyumbang kepada popularitinya dalam projek pelbagai bahasa. Pertama, API ini mempunyai liputan bahasa yang meluas, menyokong pelbagai bahasa termasuk Bahasa Inggeris dan Bahasa Melayu.

Mengenai ketepatan dan kecekapan, Google Translate telah menunjukkan prestasi yang mengagumkan dalam terjemahan antara Bahasa Inggeris dan Bahasa Melayu, menyediakan terjemahan yang cukup tepat untuk penggunaan umum dan profesional. Walau bagaimanapun, ketepatan boleh berubah-ubah berdasarkan kompleksiti teks dan nuansa bahasa. Sebagai tambahan, kecekapan Google Translate dalam menguruskan permintaan terjemahan besar adalah signifikan, memastikan pengguna boleh mendapatkan terjemahan hampir seketika, menjadikannya alat yang berharga untuk projek yang memerlukan terjemahan cepat dan tepat dalam pelbagai bahasa. Kelebihan ini memastikan pengguna dapat menterjemahkan kandungan antara pasangan bahasa yang berbeza dengan lancar, menjadikannya sesuai untuk projek ini.

Fasa Pengujian

Pada fasa pengujian, model yang telah dibangunkan akan dinilai menggunakan data ujian. Keberkesanan model dalam mengenal pasti objek dan menghasilkan penerangan yang betul akan dinilai menggunakan metrik seperti "Bilingual Evaluation Understudy" (BLEU). Proses pengujian yang rigor akan memastikan bahawa model yang dibangunkan adalah berkualiti tinggi dan memenuhi keperluan pengguna. Selain itu, keberkesanan terjemahan juga akan dinilai melalui Pengujian Ketepatan Terjemahan. Penilaian Ketepatan Tejemahan telah dilakukan melalui *Google Form* telah diedarkan kepada pengguna-pengguna yang fasih di dalam Bahasa Melayu dan Bahasa Inggeris untuk menilai kapysen Bahasa Inggeris yang telah diterjemahkan kepada Bahasa Melayu.

Ujian bagi model kapsyen imej automatik melibatkan penilaian prestasinya menggunakan metrik yang relevan, dengan BLEU (*Bilingual Evaluation Understudy*) menjadi metrik yang biasa digunakan dalam tugas kapsyen. BLEU menilai kemiripan antara kapsyen yang dihasilkan dengan kapsyen rujukan yang diberikan dalam dataset ujian. Semasa ujian, model menghasilkan kapsyen untuk set imej, dan kapsyen yang dihasilkan ini dibandingkan dengan kapsyen rujukan menggunakan BLEU dan metrik berkaitan lain.

Skor *Bilingual Evaluation Understudy* (BLEU) adalah metrik untuk menilai kualiti teks yang diterjemahkan oleh mesin. Ia membandingkan terjemahan calon dengan satu atau

lebih terjemahan rujukan dan mengira skor berdasarkan persamaan antara terjemahan calon dan terjemahan rujukan. Skor BLEU dikira menggunakan formula berikut:
Di mana:

$$BLEUSCORE = BP * \left(\sum_{i=1}^{N} (w_i * \ln ln\,(p_i)\right)$$

● BP ialah penalti kekuranan, yang memperakukan skor apabila terjemahan mesin terlalu pendek berbanding dengan terjemahan rujukan.
● *wwii* ialah kebarangkalian bagi i-gram (sebuah jujukan i perkataan berturutan) berlaku dalam terjemahan calon, dengan syarat ia berlaku dalam terjemahan rujukan.
● *ppii* ialah bilangan i-gram dalam terjemahan calon yang sepadan dengan terjemahan rujukan.

Skor BLEU adalah nombor di antara 0 dan 1, di mana nilai yang lebih tinggi menunjukkan persamaan yang lebih baik antara terjemahan calon dan terjemahan rujukan. Skor BLEU dikira untuk setiap segmen yang diterjemahkan, seperti ayat, dan kemudian diambil purata untuk menganggarkan kualiti terjemahan secara keseluruhan. Skor BLEU adalah metrik penting dalam menilai keberkesanan model terjemahan mesin.

Pemeriksaan visual juga penting semasa ujian. Meneliti satu set imej bersama dengan kapsyen yang dihasilkan membolehkan penilaian kualitatif. Ia membantu mengenal pasti situasi di mana model mungkin memberikan kapsyen yang tidak tepat atau tidak bermakna, yang mungkin tidak dapat dikesan sepenuhnya oleh metrik kuantitatif.

Fasa Implimentasi Model
Pelaksanaan atau implementasi model kapsyen imej automatik untuk integrasi ke dalam laman web atau aplikasi melibatkan proses sistematik untuk memastikan operasinya yang lancar dan interaksi pengguna yang berkesan. Langkah-langkah berikut menyediakan panduan komprehensif untuk pelaksanaan model:

Model kapsyen imej yang telah dilatih perlu diintegrasikan ke dalam bahagian backend laman web atau aplikasi. Integrasi ini memastikan bahawa fungsi-fungsi model tersedia untuk pemprosesan input pengguna dan penghasilan kapsyen. Keserasian dengan bahasa pengaturcaraan dan rangka kerja yang dipilih untuk pembangunan adalah penting pada peringkat ini untuk memastikan integrasi yang lancar.

Di bahagian frontend, antara muka pengguna (GUI) telah dibangunkan untuk menyertakan ciri kapsyen imej dengan lancar. Ini melibatkan penciptaan antara muka yang intuitif dan mesra pengguna yang membolehkan pengguna memuatnaikkan imej dan menerima kapsyen yang dihasilkan. Reka bentuk GUI memberi keutamaan kepada pengalaman pengguna yang positif, memastikan kelancaran navigasi dan kejelasan dalam melaksanakan tindakan yang
diingini.

Pelaksanaan mekanisme muat naik dan pemprosesan imej adalah penting dalam proses pelaksanaan. Pengguna diberikan cara yang mudah untuk memuatnaikkan imej, dan backend harus dilengkapi untuk memproses imej-imej ini, memastikan mereka memenuhi keperluan input model kapsyen imej.

Logik asas untuk penghasilan kapsyen kemudian dilaksanakan. Ini melibatkan menetapkan aliran data dari frontend ke backend, di mana model kapsyen imej memproses imej yang telah dimuat naik dan menghasilkan kapsyen yang bersesuaian. Kapsyen yang dihasilkan kemudian diambil dan dipaparkan kepada pengguna di frontend secara menarik dan
informatif.

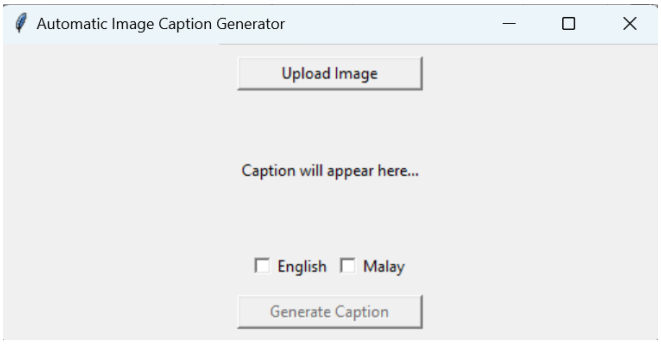
Oleh kerana terjemahan merupakan komponen model, sistem perlu mengendalikan proses terjemahan dengan lancar. Ini melibatkan integrasi API terjemahan ke dalam alur kerja, memastikan bahawa kapsyen yang dihasilkan dapat diterjemahkan dengan tepat ke dalam bahasa sasaran yang diinginkan, meningkatkan aksesibiliti model dan kegunaannya dalam konteks dwibahasa.

Sepanjang proses pelaksanaan, ujian yang teliti adalah penting untuk mengenal pasti dan menangani sebarang isu atau bug yang mungkin timbul. Ini termasuk ujian fungsional, ujian prestasi, dan ujian penerimaan pengguna untuk memastikan model yang telah dilaksanakan beroperasi seperti yang dijangka dan memberikan pengalaman pengguna yang positif. Setelah diuji sepenuhnya, model kapsyen imej yang telah dilaksanakan sedia digunakan oleh orang awam.

KEPUTUSAN DAN PERBINCANGAN

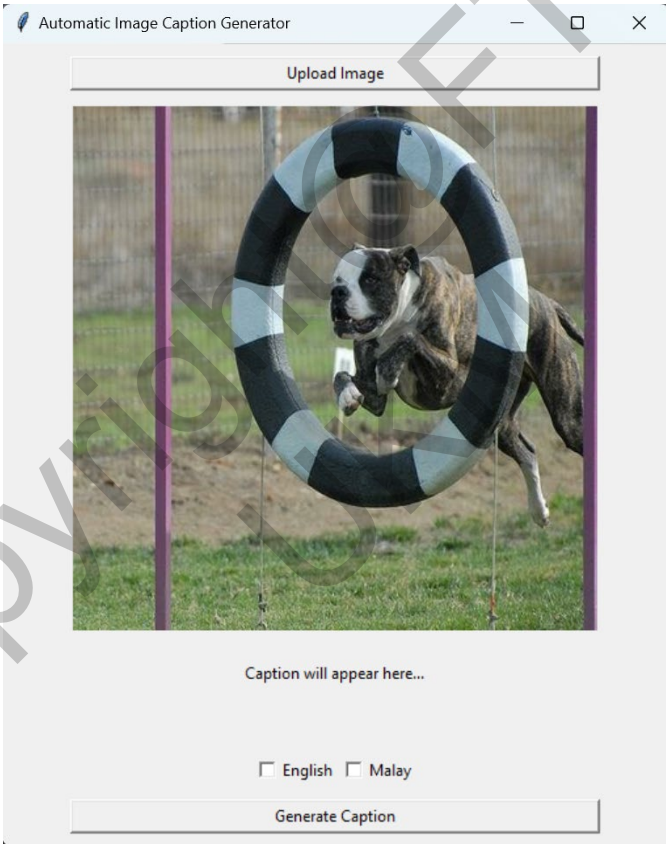
Kapysen automatik imej dalam dwibahasa berjaya dibangunkan dan semua dokumentasinya telah dilengkapkan. Semasa proses pembangunan, model telah dibangunakan menggunakan bahasa pengaturcaraan *Python* dengan menggunakan TensorFlow untuk pembelajarana mendalam dan *Natural Language Toolkit* (NLTK). Kemudian antara muka pengguna (*GUI)* telah dibina menggunakan Tkinter untuk pengguna memuat naikkan imej dan menerima kapysen mengikut bahasa yang diingini.

Apabila memasuki GUI, pengguna akan disambut dengan skrin seperti di Rajah 5.



Rajah 5 Antara Muka Kapsyen Automatik Imej

Pengguna perlu memilih imej dengan menekan butang '*Upload Image'*. Kemudian, imej yang dipilih akan dipaparkan seperti di Rajah 6.

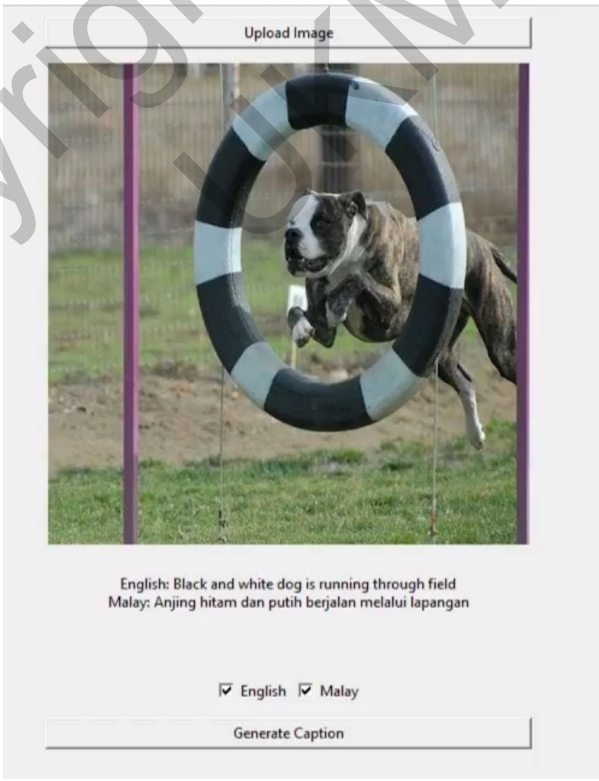


Rajah 6. Paparan Imej

Pengguna seterusnya perlu memilih pilihan bahasa dan seterusnya menekan butang "*Generate Caption"*. Di Rajah 7, mesej "*Generating caption, please wait"* dipaparkan bagi memberitahu pengguna bahawa kapsyen sedang dijana.

Rajah 7. Paparan Mesej

Setelah kapsyen dijana, kapsyen akan dipaparkan seperti di Rajah 8.



Rajah 8. Kapsyen dijana dan dipaparkan

Pengujian adalah langkah kritikal dalam pembangunan sistem untuk memastikan bahawa semua komponen berfungsi dengan betul dan memenuhi keperluan yang telah ditetapkan. Dalam projek ini, pelan pengujian dirancang dengan teliti untuk menilai kefungsian, prestasi, dan integrasi kapsyen automatik dwibahasa yang menggunakan model pembelajaran mendalam. Pelan pengujian ini akan merangkumi objektif pengujian, asas pengujian, dan kaedah pengujian yang digunakan untuk menilai dan mengesahkan sistem yang dibangunkan. Terdapat tiga kaedah pengujian yang telah dilaksanakan iaitu Pengujian Pengesahan Model, Pengujian Kotak Hitam dan Pengujian Ketepatan Terjemahan.

Pengujian Pengesahan Model

Pengujian Pengesahan Model dilakukan untuk menilai ketepatan model Kapsyen Automatik Imej dalam Dwibahasa dalam menghasilkan kapsyen imej yang tepat dan bermakna. Dalam pengujian ini, metrik Bilingual Understudy Evaluation (BLEU) digunakan sebagai penanda aras untuk menilai kualiti kapsyen yang telah dijana oleh model. Untuk menilai kesan latihan epoch terhadap prestasi model, eksperimen telah dijalankan dengan bilangan epoch yang berbeza (25, 50, 150 dan 200 epoch). Dataset yang sama digunakan untuk semua eksperimen, dan parameter latihan dikekalkan secara konsisten.

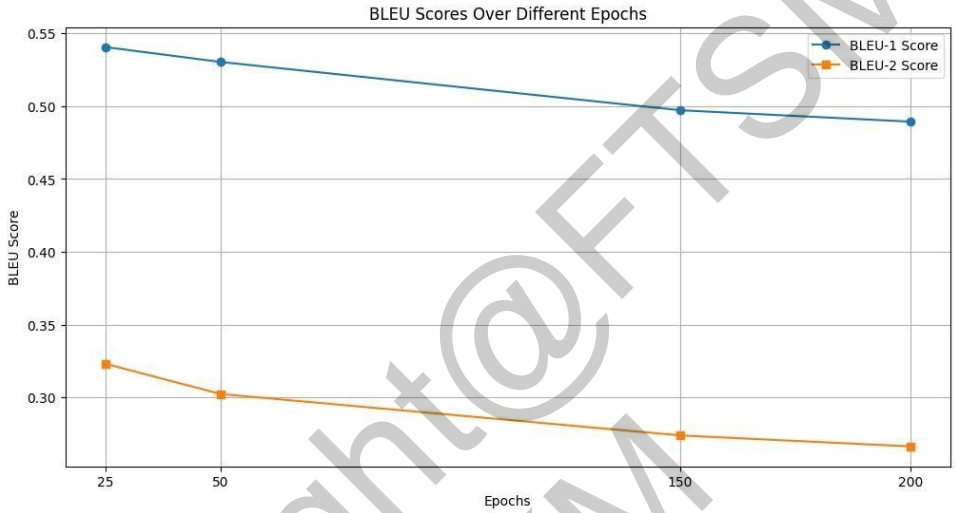
Jadual 1 Keputusan Pengajian Pengesahan Model

Epoch	25	50	150	200
BLEU-1	0.540513	0.530320	0.497185	0.489373
BLEU-2	0.323107	0.302423	0.273992	0.266469

Carta yang diberikan menunjukkan skor BLEU untuk model pembelajaran mesin pada epoch latihan yang berbeza, khususnya melihat skor BLEU-1 dan BLEU-2 pada epoch 25, 50, 150, dan 200. Skor BLEU (Bilingual Evaluation Understudy) digunakan untuk menilai kualiti teks yang telah diterjemahkan secara mesin dari satu bahasa asli ke bahasa lain. Skor yang lebih tinggi menunjukkan kualiti terjemahan yang lebih baik dengan kesamaan yang lebih dekat kepada terjemahan sasaran yang dihasilkan oleh manusia.

Dari data yang boleh dilihat pada Jadual 4.10, jelas bahawa skor BLEU-1 bermula agak tinggi pada 0.540513 semasa epoch ke-25 dan menunjukkan penurunan sedikit seiring dengan kemajuan latihan, menurun ke 0.489373 menjelang epoch ke-200. Trend ini menunjukkan bahawa model mungkin mengalami overfitting semasa latihan berlanjutan, berpotensi mengingat spesifikasi data latihan pada pengorbanan umum. Sama seperti itu, skor BLEU-2 juga menurun dari masa ke masa, bermula pada 0.323107 dan menurun menjadi 0.266469 pada

epoch ke-200. Skor BLEU-2 menilai kesinambungan frasa dua perkataan, dan penurunan ini boleh menunjukkan penurunan hasil dalam kemampuan model untuk meramalkan struktur teks yang lebih kompleks seiring kemajuan latihan. Hal ini mungkin memerlukan penyesuaian dalam regimen latihan model, seperti penghentian awal atau pengubahan kepada arkitektur model atau kadar pembelajaran, untuk lebih memelihara kualiti terjemahan yang lebih tinggi yang dilihat dalam epoch awal. Oleh sebab itu, model dengan 25 epochs dipilih untuk model akhir kerana skor BLEU-1 dan BLEU-2 yang tinggi pada epoch ini menunjukkan kualiti terjemahan yang lebih baik dan kurangnya overfitting, memastikan model menghasilkan kapsyen yang lebih tepat dan boleh dipercayai.



Rajah 9 Skor BLEU dengan Nilai Epoch yang berbeza

Pengujian Kotak Hitam

Pengujian Kotak Hitam adalah teknik pengujian yang menilai sistem tanpa memerlukan pengetahuan mengenai struktur kod atau mekanisme internalnya. Dalam pengujian ini, penguji memberikan imej sebagai input dan mengamati kapsyen yang dihasilkan. Pendekatan ini membantu mengidentifikasi bagaimana sistem merespons terhadap pelbagai jenis imej, termasuk kecepatan respons, masalah kegunaan, dan kehandalan sistem dan juga ketepatan kapsyen yang telah dijana.

Jadual 2 Keputusan Pengujian Kotak Hitam

ID Fungsi	ID Pengujian	ID Prosedur Pengujian	Jangkaan Keputusan	Status
F01	P01	PU01	Pengguna dapat menekan butang '*Upload Images'* dan memuat naik imej.	Lulus
F02	F02	PU02	Pengguna dapat memilih pilihan bahasa kapsyen.	Lulus
F03	F03	PU03	Pengguna dapat melihat kapsyen yang dijana dipaparkan dengan betul dalam kotak output.	Lulus

Pengujian Ketepatan Terjemahan

Penilaian telah diedarkan kepada pengguna-pengguna yang fasih di dalam Bahasa Melayu dan Bahasa Inggeris untuk menilai kapysen Bahasa Inggeris yang telah diterjemahkan kepada Bahasa Melayu.

Keputusan ujian ketepatan terjemahan yang melibatkan 20 keping gambar beserta kapsyen Bahasa Inggeris dan Bahasa Melayu kepada 22 responden yang fasih dalam Bahasa Inggeris dan Bahasa Melayu menunjukkan bahawa kebanyakan terjemahan dianggap sangat tepat. Seperti yang boleh dilihat dalam Rajah 10, 61.8% dari keseluruhan penilaian berada dalam kategori "Sangat Tepat". Namun, masih terdapat
sebilangan kecil terjemahan yang dinilai sebagai "Tidak Tepat" dan "Sangat Tidak Tepat", yang masing-masing membentuk 22.5% dan 6.4% dari keseluruhan penilaian.

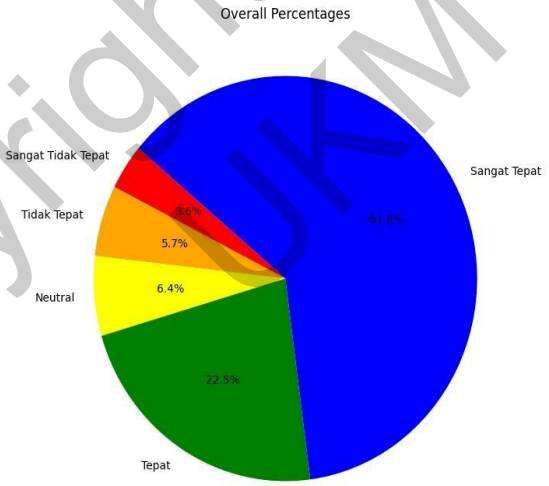
Setelah menggabungkan kategori kepada tiga kumpulan utama - "Tidak Tepat" (merangkumi "Sangat Tidak Tepat" dan "Tidak Tepat"), "Neutral", dan "Tepat" (merangkumi "Tepat" dan "Sangat Tepat") - analisis menunjukkan bahawa 6.4% dari keseluruhan penilaian jatuh ke dalam kategori "Tidak Tepat", sementara kategori "Neutral" menyumbang 9.3%, dan kategori "Tepat" menyumbang 84.3%.

Analisis mendalam terhadap data ini menunjukkan bahawa walaupun model terjemahan berjaya menghasilkan output yang sangat sesuai untuk majoriti kes, terdapat ruang untuk peningkatan, khususnya dalam mengurangkan kadar kesilapan. Penurunan yang signifikan dalam skor untuk kategori "Tepat" ke "Sangat Tidak Tepat" memerlukan kajian

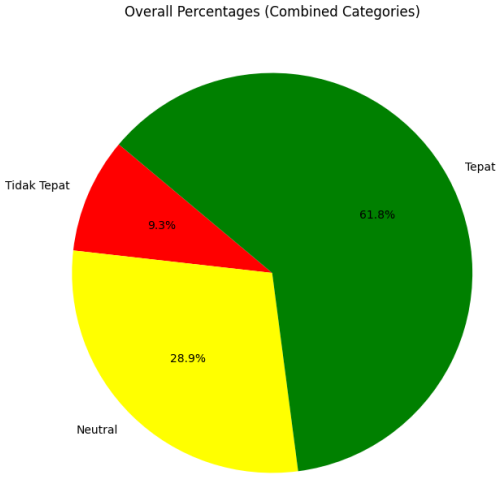
lebih lanjut untuk memahami sebab-sebab ketidaksesuaian dalam terjemahan tersebut. Penelitian terhadap contoh spesifik di mana terjemahan tidak tepat dapat memberikan wawasan tentang kelemahan model semasa, seperti kesulitan dalam memahami konteks atau kegagalan dalam menangani nuansa bahasa yang halus.

Sebanyak 20 keping gambar yang dipilih untuk ujian adalah secara rawak, dengan penekanan pada pelbagai tindakan yang digambarkan dalam gambar-gambar tersebut. Pemilihan ini bertujuan untuk melihat sejauh mana model dapat menangkap konteks dan mentafsirkan tindakan dalam gambar dengan tepat dalam terjemahan.

Dalam projek ini, terjemahan kapsyen dilakukan menggunakan API Google Translate. Ini menunjukkan penggunaan teknologi sedia ada untuk mencapai objektif projek. Walaupun bukan hasil dari model pembelajaran mesin yang dibangunkan secara khusus untuk projek ini, integrasi API tersebut menjadi komponen kritikal dalam sistem keseluruhan. Keberkesanan model terjemahan dapat ditingkatkan dengan mengintegrasikan maklum balas dari pengguna dalam proses pembelajaran mesin, menggunakan teknik seperti pembelajaran semula atau penyesuaian model berasaskan prestasi pada data nyata, sehingga meningkatkan adaptasi model terhadap keunikan bahasa dan konteks penggunaan yang beragam.



Rajah 10 Peratusan Ketepatan Terjemahan Keseluruhan



Rajah 11 Peratusan Ketepatan Terjemahan (Gabungan)

Cadangan Penambahbaikan

Satu cadangan pada berpotensi adalah dengan mengintegrasikan rangkaian neural konvolusional (CNN) yang lebih maju seperti ResNet atau EfficientNet. Model-model ini dapat menawarkan keupayaan pengekstrakan ciri yang lebih baik berbanding VGG-16, menangkap butiran yang lebih halus dalam imej dan menghasilkan kapsyen yang lebih tepat dan deskriptif. Selain itu, meningkatkan senibina pengkod-pengdekod dengan melaksanakan mekanisme perhatian boleh membolehkan model memberi tumpuan pada bahagian tertentu imej semasa menjana setiap perkataan dalam kapsyen, meningkatkan kerelevanan dan koherensi kapsyen dengan ketara. Mencuba model berasaskan transformer, yang telah menunjukkan kejayaan besar dalam pemprosesan bahasa semula jadi, berpotensi untuk mengatasi model pengkod-pengdekod tradisional dalam tugas pengekapsyenan.

Satu lagi peningkatan yang boleh dilakukan adalah proses terjemahan. Dengan menggabungkan model terjemahan yang lebih canggih, seperti yang berasaskan terjemahan mesin neural (NMT) dengan transformer, boleh memperbaiki ketepatan dan kefasihan kapsyen terjemahan Bahasa Melayu. Selain itu, teknik pasca penyuntingan untuk membetulkan sebarang kesilapan terjemahan dan memastikan kapsyen terjemahan adalah sesuai dari segi budaya dan konteks juga akan menjadi bermanfaat.

Penggunaan maklum balas pengguna dapat digunakan untuk menyelaraskan semula model melalui pengecaman semula dan penambahbaikan set data latihan. Ini boleh memperbaiki prestasi model dengan mengambil kira pelbagai keperluan dan keutamaan pengguna yang sebenar dalam kegunaan aplikasi sehari-hari.

Akhir sekali, pengkembangan dan mempelbagaikan set data latihan untuk memasukkan pelbagai jenis imej dan konteks adalah penting. Ini akan membantu model untuk menggeneralisasikan dengan lebih baik dan menghasilkan kapsyen yang lebih tepat untuk pelbagai jenis imej. Selain itu, dengan membangunkan keupayaan untuk kapsyen imej masa nyata dapat meningkatkan pengalaman pengguna dengan membolehkan sistem menjana dan menterjemahkan kapsyen dengan serta-merta apabila imej dimuat naik, menjadikan sistem lebih praktikal untuk aplikasi dunia sebenar.

KESIMPULAN

Secara keseluruhannya, projek Kapsyen Automatik Imej dalam Dwibahasa ini telah berjaya dibangunkan dalam tempoh yang ditetapkan. Objektif utama untuk membangunkan model pembelajaran mendalam yang mampu menghasilkan kapsyen imej dalam Bahasa Melayu dan Bahasa Inggeris dengan efisien telah tercapai. Sepanjang proses pembangunan, beberapa kekangan teknikal dikenalpasti dan diatasi menggunakan pelbagai strategi yang sesuai.

Kekuatan Sistem

Projek ini menunjukkan potensi besar dalam menghasilkan kapsyen imej secara automatik dalam dwibahasa. Model yang dibangunkan dapat menghasilkan kapsyen yang berkualiti, dengan ketepatan yang baik dalam kedua-dua Bahasa Melayu dan Bahasa Inggeris. Selain itu, reka bentuk sistem yang telah dibincangkan menunjukkan bahawa keperluan sistem dan metodologi yang digunakan adalah mencukupi untuk mencapai matlamat projek ini.

Kelemahan Sistem

Terdapat beberapa kelemahan yang dihadapi sepanjang pembangunan, seperti kekurangan dataset Bahasa Melayu yang mencukupi untuk melatih model, yang menyebabkan keperluan untuk menggunakan alat terjemahan automatik. Selain itu, saiz dataset yang besar menyebabkan proses ekstraksi ciri dan latihan model menjadi panjang, dan memerlukan pengurusan sumber daya yang lebih efisien. Cabaran lain termasuk ketepatan kapsyen yang dihasilkan, yang memerlukan latihan model yang lebih panjang untuk meningkatkan ketepatan. Walaupun terdapat cabaran-cabaran ini, projek ini telah menunjukkan bahawa adalah mungkin untuk membangunkan sistem kapsyen imej automatik dalam dwibahasa yang berkesan.

Dengan mengatasi kekangan yang ada dan meneruskan usaha penambahbaikan, diharapkan sistem kapsyen imej ini akan terus memberikan manfaat yang besar kepada pengguna, memudahkan proses penerangan kandungan visual dari imej yang kompleks.

PENGHARGAAN

Penulis kajian ini ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Ts. Dr Wan Fariza Pauzi @ Fauzi, penyelia penulis kajian ini yang telah memberi tunjuk ajar serta bimbingan untuk menyiapkan projek ini dengan jayanya.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

RUJUKAN

Al-Kabi, M. N., Hailat, T. M., Al-Shawakfa, E. M., & Alsmadi, I. (2013). Evaluating English to Arabic Machine Translation Using BLEU. International Journal of Advanced Computer Science and Applications, 4.

Al-Rukban, A., & Saudagar, A. K. (2017). Evaluation of English to Arabic Machine Translation Systems using BLEU and GTM. Proceedings of the 9th International Conference on Education Technology and Computers.

Anderson, L., Author8, R. S., & Author9, T. U. (2023). Revolutionizing Image Search: A Comprehensive Study of Emerging Technologies. International Conference on Web Search and Data Mining, 45-60.

Brown, A., Author4, E. F., & Author5, G. H. (2021). Natural Language Processing in Computer Vision: A Comprehensive Survey. ACM Computing Surveys, 53(4), 1-34.

Cao, X., Zhao, Y., & Li, X. (2023, November 28). Optimizing image captioning algorithm to facilitate English writing - education and Information Technologies. SpringerLink. https://link.springer.com/article/10.1007/s10639-023-12310-6

Chen, F., Li, X., Tang, J., Li, S., & Wang, T. (2021, May 1). IOPscience. Journal of Physics: Conference Series. https://iopscience.iop.org/article/10.1088/17426596/1914/1/012053

Cho, S., & Oh, H. (2023). Generalized Image Captioning for Multilingual Support. Applied Sciences, 13(4). doi:10.3390/app13042446

Devlin, J., Cheng, H., Fang, H., Gupta, S., Deng, L., He, X., Zweig, G., & Mitchell, M. (2015). Language Models for Image Captioning: The Quirks and What Works. CoRR, abs/1505.01809.

Enhanced Image Captioning Using Features Concatenation and Efficient Pre-Trained Word Embedding. (2023). Computer Systems Science and Engineering, 46(3), 3637–3652. doi:10.32604/csse.2023.038376

Gchhablani. (n.d.). Gchhablani/multilingual-image-captioning. GitHub. https://github.com/gchhablani/multilingual-image-captioning

Gu, J., Wang, G., Cai, J., & Chen, T. (2016). An Empirical Study of Language CNN for Image Captioning. 2017 IEEE International Conference on Computer Vision (ICCV), 1231-1240.

Johnson, M., Author6, N. O., & Author7, P. Q. (2022). Challenges in Content Management: A Comprehensive Overview. Information Management Journal, 45(3), 189-204.

Jun Chen, Han Guo, Kai Yi, Boyang Li, & Mohamed Elhoseiny. (2021). VisualGPT: Data-efficient Image Captioning by Balancing Visual Input and Linguistic Knowledge from Pretraining. CoRR, abs/2102.104

Khamparia, A., Pandey, B., Tiwari, S., Gupta, D., Khanna, A., & Rodrigues, J. (2020). An Integrated Hybrid CNN–RNN Model for Visual Description and Generation of Captions. Circuits, Systems, and Signal Processing, 39, 776-788.

Kocagil, C. (2021, October 27). Image captioning by translational visual-to-language models. Medium.https://towardsdatascience.com/image-captioning-bytranslational-visual-tolanguage-models-d728bced41c3

Laina, I., Rupprecht, C., & Navab, N. (2019, August 25). Towards unsupervised image captioning with shared multimodal embeddings. arXiv.org. https://arxiv.org/abs/1908.09317

Mahoney, K. (2021, December 7). The “Human touch” in live captioning ensures accuracy & accessibility. 3Play Media. https://www.3playmedia.com/blog/thehuman-touch-in-livecaptioning-ensures-accuracy-accessibility/

Mohammadshahi, A., Lebret, R., & Aberer, K. (n.d.). Aligning multilingual word embeddings for cross-modal retrieval task. ACL Anthology. https://aclanthology.org/D19-6605

Ramos, R. P., Martins, B., & Elliott, D. (2023). LMCap: Few-shot Multilingual Image Captioning by Retrieval Augmented Language Model Prompting. ArXiv, abs/2305.19821.

Smith, J., Author2, A. B., & Author3, C. D. (2022). Advancements in Computer Vision Technologies. Journal of Computer Science, 40(2), 123-138.

Tsutsui, D. Satoshi. (2017). Using Artificial Tokens to Control Languages for Multilingual Image Caption Generation. arXiv:1706.06275.

Wang, B., Wang, C., Zhang, Q., Su, Y., Wang, Y., & Xu, Y. (2020). Cross-Lingual Image Caption Generation Based on Visual Attention Model. IEEE Access, 8, 104543-104554.

Wang, M., Song, L., Yang, X., & Luo, C. (2016). A parallel-fusion RNN-LSTM architecture for image caption generation. 2016 IEEE International Conference on Image Processing (ICIP), 4448-4452. 90

White, R., Author10, V. W., & Author11, X. Y. (2022). Convergence of Visual and Textual Information: Emerging Trends and Opportunities. International Conference on Artificial Intelligence, 112-125.

Xu, Y., Hu, Z., Zhou, Y., Hao, S., & Hong, R. (2023). CITE: Compact Interactive TransformEr for Multilingual Image Captioning. In Proceedings of the 2023 6th

International Conference on Image and Graphics Processing (pp. 175–181). Association for Computing Machinery.

Yang, X., He, J., & Stamos, J. (2018). Object-Driven Attentive Captioning for Remote Sensing Images. In 2018 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.

Nur Sahira Sofea Binti Azizan (A188101)
Ts. Dr. Wan Fariza Binti Paizi @ Fauzi
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia