Pembangkit Teks Otomatis Citra Geologi Bebatuan Dengan Deteksi Relasi *Semantic Attention* (SemAtt) Antar Objek

RINGKASAN DISERTASI

Agus Nursikuwagus
NIM: 33218302
(Program Studi Doktor Teknik Elektro dan Informatika)

Institut Teknologi Bandung
Agustus 2023

Pembangkit Teks Otomatis Citra Geologi Bebatuan Dengan Deteksi Relasi *Semantic Attention* (SemAtt) Antar Objek

Disertasi ini dipertahankan pada Sidang Terbuka Sekolah Pascasarjana sebagai salah satu syarat untuk memperoleh gelar Doktor Institut Teknologi Bandung

Agustus 2023

Agus Nursikuwagus
NIM: 33218302
(Program Studi Doktor Teknik Elektro dan Informatika)

Promotor : Dr. Ir. Rinaldi Munir, MT
Ko-Promotor : Dr. Masayu L. K, ST.,MT

Institut Teknologi Bandung
Agustus 2023

Pembangkit Teks Otomatis Citra Geologi Bebatuan Dengan Deteksi Relasi *Semantic Attention* (SemAtt) Antar Objek

Agus Nursikuwagus
NIM 33218302

1. Latar Belakang

Pengamatan geologi bebatuan merupakan penelitian lapangan oleh seorang ahli geologi. Salah satu tugasnya adalah mengambil foto dan memberikan deskripsi citra bebatuan. Setiap citra dipasangkan dengan deskripsinya yang disebut dengan *caption*. Setiap deskripsi mencantumkan nama batuan (klasifikasi), warna, dan pola batuan (Joko et al., 2017). Kegiatan ini merupakan pekerjaan penting karena informasi yang dibuat akan dipergunakan dalam pengambilan keputusan. Pengamatan, pengambilan foto, dan deskripsi merupakan suatu kegiatan yang sistematis dari seorang ahli geologi dalam memberikan deskripsi citra bebatuan. Pada saat pengamatan bebatuan, seorang ahli geologi kerap kali menemukan batuan sejenis dengan klasifikasi, warna dan pola yang sama. Proses deskripsi seringkali menuliskan deskripsi yang sama pada citra yang diamati. Hal ini dapat memicu pekerjaan penulisan deskripsi berulang, dan menyebabkan pekerjaan berlangsung lambat.

Pemberian deskripsi citra bebatuan dalam bidang komputasi dapat didekati dengan bidang *natural language processing (NLP)* dan *computer vision (CV)*. Rangkaian kerja tersebut merupakan proses produksi kalimat yang mengandalkan pengamatan/*vision*. Pemetaan nama dan warna bebatuan terhadap citra yang diamati oleh ahli geologi menjadi faktor kunci dalam pemerian citra. Keahlian *vision* dari ahli geologi merupakan rekognisi yang dapat menginterpretasikan citra bebatuan. Rekognisi tersebut diformulasi menjadi suatu pengenalan atau pengklasifikasian citra pada teknik *vision*. Kemudian hasil formulasi rekognisi citra digabungkan dengan kumpulan kata yang merupakan representasi citra membentuk suatu formulasi yang dapat diintepresikan sebagai suatu model. Penggabungan model *vision* dan *natural language processing* menjadikan suatu disiplin ilmu yang dikenal dengan istilah yaitu *captioning* citra (Karpathy and Fei-Fei, 2015). *Captioning* yang dikerjakan ahli geologi dapat dimodelkan dengan menerapkan *computer vision* dan *natural language processing*. *Captioning* untuk citra geologi bebatuan merupakan arsitekur gabungan *convolutional neural network* (CNN) dan *long short-term memory* (LSTM). *Captioning* ini dipergunakan untuk mengidentifikasi citra geologi bebatuan, sehingga dapat mereduksi *caption* berulang pada citra yang memiliki kemiripan bebatuan geologi.

Sebagian besar penelitian telah mengusung beberapa konsep model *captioning* citra yang menitikberatkan pada model *encoder* dan *decoder* (R. Li et al., 2020). Arsitektur *encoder* dan *decoder* yang dibuat dapat menghasilkan teks yang mendekati deskripsi seseorang. Pada arsitektur *encoder*, *task* yang dilakukan yaitu *ekstraksi* citra. Arsitekturnya banyak menggunakan metode CNN. Krizhevsky dkk, (2012) mengusulkan CNN, klasifikasi *ImageNet*, serta memberikan kontribusi untuk model ekstraksi citra (Krizhevsky et al., 2012). YOLO (Redmon and Farhadi, 2018), GoogLeNet (Szegedy et al., 2015), VGGNet (Simonyan and Zisserman, 2014), Resnet (Kaiming et al., 2016), dan InceptionV3 (Bhatia et al., 2019; Chun et al., 2022) merupakan model untuk rekognisi citra. Pada arsitektur *decoder*, *task* yang dilakukan adalah membangkitkan kata yang memiliki relasi dengan area citra. Setiap peta fitur hasil dari *task encoder* akan dipasangkan dengan *word embedding* sebagai masukan untuk *decoder* (Bahdanau et al., 2015; S. He et al., 2020; Luong et al., 2015; Vaswani et al., 2017a). *Decoder* yang dikembangkan untuk *captioning* citra seperti RNN dan LSTM (Huang et al., 2016; N. Li and Chen, 2018; Szegedy et al., 2014). Selain model tersebut, dikembangkan juga model *Transformers* untuk menghasilkan kalimat yang dapat mewakili objek yang ada di citra (Lee et al., 2020; G. Li et al., 2019).

Karpathy dkk, (2015) memperkenalkan konsep *captioning* citra berdasarkan identifikasi objek yang terdapat pada citra MS COCO (Karpathy and Fei-Fei, 2015). Karpathy berhasil berkontribusi dalam hal identifikasi objek pada area citra serta mengusulkan model *captioning* dengan arsitektur CNN dan *Bidirectional Recurrent Neural Network* (BRNN) untuk citra *MS COCO* dan *FLICKR* (Bhatia et al., 2019; Karpathy and Fei-Fei, 2015; Tanti and Camilleri, 2016).

Identifikasi objek utama dan arsitektur model sering dijadikan kontribusi dalam penelitian *captioning* (S. He et al., 2020; Lee et al., 2020; Su et al., 2019). Bila diperhatikan mengenai objek pada citra, sesungguhnya ada dua jenis yang dapat diamati, yaitu objek latar depan dan objek latar belakang. Objek latar depan dapat diidentifikasi sebagai objek utama dengan memperhatikan kelas objek, seperti mobil, pria, wanita, jalan, tangga, rumput, burung, hewan, dan lainnya (Karpathy and Fei-Fei, 2015).

Objek latar belakang adalah area yang terletak di belakang objek utama, seperti tembok, pekarangan, bebatuan, awan, hamparan padang rumput, dan langit seperti pada Gambar 1(b). Fokus objek latar belakang bukan pada objek palu Gambar 1(b) tetapi objek yang ada dibelakang palu tersebut. Beberapa studi penelitian telah memperkenalkan model *captioning* citra yang menekankan pada objek latar depan seperti pada Gambar 1(a) (Chun et al., 2022). Bila diperhatikan, bahwa objek bebatuan pada citra geologi merupakan objek yang menjadi latar belakang dari objek utama. Objek bebatuan inilah yang dijadikan fokus identifikasi untuk diberikan

caption (Simonyan and Zisserman, 2014) Konsep identifikasi objek latar belakang dengan *semantic attention* secara prinsip memiliki kerja yang mirip dengan identifikasi objek utama. Objek citra geologi bebatuan lebih memperhatikan objek latar belakang, hal inilah yang menjadi sasaran penelitian bagaimana identifikasi objek latar belakang dapat dikenal dengan tepat. *Semantic attention* yang diusulkan lebih diarahkan bagaimana hasil ekstraksi fitur citra dapat berelasi dengan fitur teks sehingga dapat memprediksi caption yang mendekati caption referensi. Studi yang dilakukan oleh (Chun et al., 2022) mengenai *captioning* citra yang mendeskripsikan mengenai objek yang mirip yang dilakukan pada penelitian ini. Pada Gambar 1(a) merupakan hasil *captioning* citra untuk mendekripsikan citra korosi dari suatu jembatan.

(a) (b)

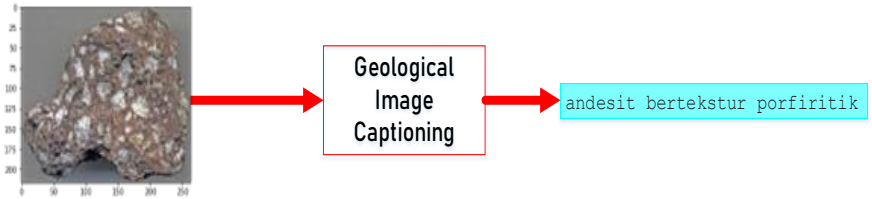
Gambar 1 Ilustrasi (a) "*a dirty old room*" (Chun et al., 2022) , (b) *caption* "Batupasir kompak retak-retak kelabu"

Eksperimen terdahulu yang menggunakan dataset citra bebatuan geologi dan pengembangan model VGG16-LSTM-*OneHotVector* hanya menunjukkan skor BLEU-1=0,5321, BLEU-1=0,4890, BLEU-1=0,4898, dan BLEU-1=0,4381 (Vinyals and Toshev, 2015). Objek yang terdekteksi pada citra bebatuan seperti manusia, berbagai benda, tanaman, dan objek yang terdefinisi pada ImageNet (Karpathy and Fei-Fei, 2015; Krizhevsky et al., 2012; Vinyals and Toshev, 2015). Objek-objek seperti bebatuan, hamparan pasir, hamparan tanah, sungai, dan pepohonan belum teridentifikasi pada model ini. Peluang penelitian dalam topik *captioning* yang memperhatikan objek latar belakang masih dimungkinkan untuk diusulkan. Perkembangan metode CNN sebagai *ekstraksi citra* dan LSTM untuk pembangkit teks, telah mendorong berkembangnya arsitektur *captioning* citra. Kata yang dihasilkan oleh model LSTM, kemudian disusun dengan menggunakan algoritma Beam Search atau Greedy Search (Karpathy and Fei-Fei, 2015; Vinyals et al., 2017; Vinyals and Toshev, 2015).

Perkembangan konsep lainnya adalah seperti *attention* dan *Transformers* (Bahdanau et al., 2015; Luong et al., 2015; Vaswani et al., 2017a). Metode *decoder* yang memproses masukan secara paralel dibandingkan sekuensial (Bahdanau et al., 2015; Luong et al., 2015; Vaswani et al., 2017a). Konsep lainnya mengenai *captioning* citra, dikembangkan juga dengan *visual attention* dan *semantic attention* (Bahdanau et al., 2015; L. Li et al., 2017; Quanzeng et al., 2016). Konsep *visual attention* dan *semantic attention* yang diusung Quanzeng dkk, (2016) mengedepankan perbaikan pada sisi arsitektur *ekstraksi citra* dengan cara menyusun layer CNN seperti arsitektur GoogleNet (Szegedy et al., 2015). Layer CNN ini bekerja sebagai *task visual attention* untuk menangani rekognisi citra. Konsep *semantic attention* merupakan *task* yang mengusung perbaikan pada bagian *decoder*. *Task* ini bekerja dengan menambahkan fungsi *attention* yang terdiri dari *Query* (Q), *Keys* (K), dan *Value* (V) (Bahdanau et al., 2015; Luong et al., 2015; Vaswani et al., 2017a).

Konsep *separable* CNN yang diusung oleh Chollet dkk, (2017) dan dikenal dengan nama *xception* dijadikan model rekognisi fitur citra. Model separable CNN digabungkan dengan model bahasa *Transformers* yang diusung oleh Vaswani dkk, (2017) dijadikan model usulan *captioning* citra bebatuan. Identifikasi objek bebatuan dengan *semantic attention* bebatuan, dijadikan fokus utama dalam mengeksplorasi model *captioning* citra. *Attention* yang diusulkan merupakan model *Transformers* yang terdiri dari dua bagian yaitu *Multihead Attention* sebagai *encoder* dan bagian *decoder*. Ketepatan, dan ketersedian kata menjadi target dalam menghasilkan *caption* yang sesuai dengan *caption* referensi dari ahli geologi. Setiap model captioning citra selalu diukur mengenai keberhasilan dalam memproduksi caption. Salah satu metrik yang digunakan adalah skor *Bilingual Evaluation Understudy* (BLEU) sebagai pengevaluasi model atas keberhasilan *caption* yang diproduksi (Papineni dkk., 2002).

Gambar 2 menunjukkan masalah penelitian disertasi yang dilakukan. Kalimat yang dihasilkan oleh ahli geologi dapat berupa kata nama batuan, dan ditambah dengan kata seperti warna (Joko et al., 2017).



Gambar 2 Masalah penelitian disertasi tentang pemberian *caption* geologi.

Berdasarkan paparan latar belakang, maka masalah penelitian adalah bagaimana model arsitektur *captioning* citra geologi untuk pembangkit *caption* citra bebatuan. Beberapa turunan dari *research question* yang terkait dengan penelitian adalah:

a. Permasalahan pertama yang diperoleh adalah bagaimana model arsitektur ekstraksi fitur (*encoder*) citra bebatuan dengan struktur jaringan reguler ataupun *separable* CNN (Chun et al., 2022; Karpathy and Fei-Fei, 2015; Lecun et al., 1998; Vinyals and Toshev, 2015). Permasalahan pada CNN yaitu bagaimana arsitektur regular CNN dan *separable* CNN dapat mendeteksi fitur edge sehingga dapat dikenali sebagai semantik dari fitur teksnya.
b. Konsep LSTM, *attention*, atau *Transformers* sebagai model *decoder* yang dapat mendorong menghasilkan *caption*, dijadikan fondasi untuk membuat model *decoder* (Bahdanau et al., 2015; Luong et al., 2015; Vaswani et al., 2017b). Permasalahan yang kedua adalah bagaimana model arsitektur pembangkit kata (*decoder*) dengan arsitektur *recurrent* atau *transformers* untuk memperoleh *caption* yang mendekati *caption* referensi. Penggalian aspek teks yang akan dijadikan prediksi kata menjadi penting ketika dilakukan operasi penggabungan antara fitur citra dan fitur teks.
c. Kegiatan yang dilakukan seperti membuat model gabungan *captioning* sangat perlu untuk diukur keberhasilannya. Hal ini menjadi masalah yang ketiga yaitu bagaimana melakukan pengukuran kinerja presisi *caption* yang diperoleh dari model *captioning* bebatuan (Chun et al., 2022; K. He et al., 2018; Karpathy and Fei-Fei, 2015; Vinyals and Toshev, 2015).

2. Tujuan dan Sasaran Penelitian

Pemaparan pada subbab latar belakang merupakan rangkaian dari gambaran permasalahan atau *task* yang dikerjakan pada saat ini. Bila didasari dari masalah penelitian, maka tujuan penelitian yang diajukan adalah sebagai berikut :

a. Mendapatkan model arsitektur ekstraksi fitur citra dengan memanfaatkan metode CNN yang memperhatikan relasi fitur antar objek. Persoalan penyelesaian tujuan ini adalah adalah menghasilkan arsitektur ekstraksi citra bebatuan geologi dengan regular atau *separable* CNN yang disebut sebagai *encoder* (Chollet, 2017; Lecun et al., 1998; Simonyan and Zisserman, 2014).
b. Mendapatkan model arsitektur pembangkit kata (*decoder*) untuk membangkitkan *caption* yang mendekati referensi. Persoalan ini dijadikan sasaran penyelesaian masalah untuk mendapat prediksi kata dan membangkitkan semantik bebatuan dengan bersandar pada metode LSTM (N. Li and Chen, 2017; Zhu et al., 2016), *Attention* (Bahdanau et al., 2015; Luong et al., 2015), ataupun metode *transformers* (Vaswani et al., 2017a).
c. Mendapatkan hasil tingkat kepresisian *caption* yang melebihi model *baseline* dengan pengukuran kepresisian *caption* menggunakan skor BLEU (Papineni et al., 2002) .

Untuk mencapai tujuan tersebut, maka perlu adanya sasaran penelitian agar penelitian tepat sasaran dan mendapatkan hasil sesuai tujuan. Batasan penelitian yang diberikan adalah sebagai berikut:

a. Dataset citra bebatuan dengan ukuran 224x224 dan 299x299 piksel dan *caption* referensi citra sebanyak lima kalimat untuk satu citra dalam bentuk dataset geologi bebatuan.
b. Masukan nilai vektor untuk model *decoder* merupakan nilai fitur vektor dari *word embedding word2vec* dan *Onehotvector* beserta nilai fitur vektor dari peta fitu.
c. Metode *CNN* sebagai *encoder* yang digunakan merupakan arsitektur regular CNN, atau *separable* CNN
d. Model bahasa sebagai *decoder* yang digunakan adalah *LSTM, LSTM+Attention*, serta *Transformers* digunakan sebagai pembangkit kata.
e. Strukturisasi *caption* bebatuan menggunakan algoritma *Greedy Search* dan *Beam Search*.
f. Evaluasi model captioning *citra* dilakukan dengan menggunakan skor *BLEU* yang mengevaluasi *caption* referensi dengan *caption* hasil.

3. Metode Penelitian

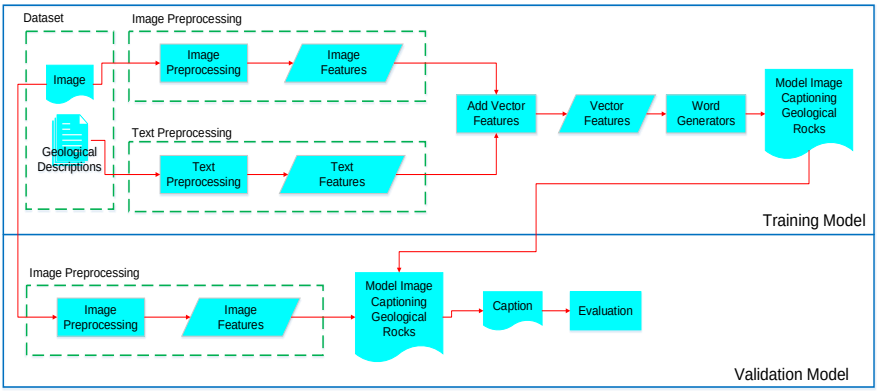
Usulan penelitian merupakan rangkaian metode yang dikerjakan selama penelitian disertasi berlangsung. Langkah operasi ini perlu digambarkan dalam suatu bentuk diagram agar dapat dipahami arah dan luaran dari penelitian tersebut. Berdasarkan model pembelajaran mesin yang telah disepakati, maka model ini dibagi menjadi dua bagian yaitu model pelatihan dan model validasi..

Blok rekognisi citra yang didukung oleh CNN dirancang untuk mendapatkan ektraksi fitur yang dapat mewakili karakteristik dari bebatuan. Usulan yang dirancang adalah berupa *regular* CNN (Szegedy et al., 2014, 2015) dan *separable* CNN (Bhatia et al., 2019; Chollet, 2017). Target yang ingin diperoleh dari blok CNN ini adalah mendapatkan rekognisi citra geologi yang sarat dengan warna yang hampir mirip. Waktu dan akurasi menjadi pertimbangan pada blok rekognisi citra ini, operasi *separable* CNN memberikan kinerja waktu yang lebih baik dari pada regular CNN (Chollet, 2017). Penentuan metode pembangkit kata dengan kepresisian hasil caption menjadi sasaran keberhasilan *caption*. Blok rokognisi citra (*encoder*) dan blok pembangkit kata (*decoder*) dijadikan target penelitian untuk keberhasilan memberikan *caption* pada citra geologi bebatuan.

Berdasarkan domain citra yang spesifik seperti citra geologi bebatuan dan penelitian dalam domain klasifikasi bebatuan geologi masih terbuka untuk penelitian *captioning* citra. Hal ini belum dikerjakan sebagai penelitian *captioning* citra. Kontribusi yang diusulkan merupakan rancangan arsitektur yang berdasarkan metode *ensemble* (gabungan) yang terdiri dari blok CNN dan blok pembangkit kata.

Perbedaan dengan penelitian Karpathy terletak pada arsitektur dan *encoding* kata sebagai masukan untuk *word embedding*. Karpathy menggunakan metode vektor *onehot*, pada usulan ini menggunakan *word2vec*. Berdasarkan usulan metode maka target yang diperoleh adalah *caption* geologi bebatuan yang mendekati *referensi* ahli geologi.

Tahap Penelitian satu. Tahap ini melakukan eksperimen pembangkit caption citra geologi dengan menggunakan model *baseline* yang berkembang dan dijadikan target hasil dari berbagai penelitian. Selain melakukan eksperimen pada tahap ini, penelitian juga melakukan pengumpulan data yang berasal dari sumber yaitu ahli geologi (Joko et al., 2017). Eksperimen yang dilakukan adalah mengujikan model *captioning baseline* (Karpathy and Fei-Fei, 2015; Szegedy et al., 2014, 2016; Vinyals and Toshev, 2015). *Baseline* yang digunakan yaitu VGG16, ResNet50, dan InceptionV3 yang digabungkan dengan LSTM (Wang et al., 2016) sebagai model *decoder* pembangkit kata. *Word embedding* yang digunakan adalah *onehotvector*. Eksperimen ini dilakukan untuk mendapatkan arsitektur model *captioning* yang terbaik dari model *baseline captioning*. Hasil dari pembangkitkan kata maka akan dilakukan rekonstruksi *caption* dengan menggunakan algoritm *Beam search* atau *greedy search*. Kemudian *caption* akan dievaluasi dengan menggunakan skor *BLEU* untuk melihat tingkat ketepatan *caption*.



Gambar 3. Metode Penelitian

Dataset citra geologi yang digunakan adalah citra hasil penyelidikan lapangan mengenai bebatuan geologi yang tampak. Pada eksperimen yang dilakukan, dataset yang digunakan adalah sebanyak 843 citra yang sudah memiliki *caption* dari ahli geologi (Joko dkk., 2017). Dataset untuk *caption* merupakan hasil deskripsi dari ahli geologi mengenai citra yang diamati. Deskripsi ini diberikan pada citra sebanyak lima deskripsi untuk satu citra. Sehingga total deskripsi adalah 4215 deskripsi.

Anotasi terhadap caption dilakukan langsung oleh ahli geologi, pekerjaan yang dilakukan pada penelitian ini meliputi strukturisasi dataset dan pengolah mula teks. Text pengolah mula meliputi *remove punctuation* dan *lower case text*. Dataset ini tersusun atas dua buah atribut yaitu nama citra dan caption dari citra tersebut. Pada **Error! Not a valid bookmark self-reference.** merupakan contoh dataset citra geologi yang dikumpulkan berdasarkan keterangan ahli geologi.

Tahap Penelitian dua. Tahap ini merupakan keberlanjutan dari tahap pertama dengan tetap menggunakan dataset yang telah diperoleh. Tahap ini mengusulkan rekayasa dari *word embedding* yaitu dengan menggunakan *word2vec* (Mikolov et al., 2013) dengan ekstraksi citra masih menggunakan *VGG16*, *ResNet50*, dan *InceptionV3*. Tahap ini juga mengusulkan *backbone CNN* hasil dari eksperimen klasifikasi citra mengenai bebatuan geologi. Model ekstraksi citra atau *backbone CNN* hasil eksperimen baru digabungkan dengan LSTM sebagai model *decoder*. Kemudian hasil *caption* dievaluasi kembali dengan menggunakan skor *BLEU* untuk melihat perubahan hasil yang diperoleh.

Tabel 1 Contoh dataset citra geologi caption

	15.jpg#0 Singkapan batuan sedimen klastik dengan bidang perlapisan yang tidak jelas, sebagian hancur dan lapuk 15.jpg#1 Singkapan batuan sedimen klastik dan batulumpur 15.jpg#2 Singkapan Konglomerat dan batulumpur berwarna kelabu 15.jpg#3 Singkapan batuan sedimen konglomerat 15.jpg#4 Singkapan Konglomerat
	142.jpg#0 Intrusi Breksi berwarna abu-abu 142.jpg#1 breksi abu-abu 142.jpg#2 breksi dengan batu kristal 142.jpg#3 Intrusi breksi dengan batu kristal 142.jpg#4 intrusi breksi abu-abu dengan batu kristal

Tahap Penelitian tiga. Tahap ini adalah melakukan rekayasa kembali dari *backbone CNN* yang telah diperoleh sebagai pengekstraksi citra. Rekayasa yang dilakukan adalah pada bagian *semantic attention* yaitu model pembangkit kata LSTM. Usulan yang dilakukan adalah dengan menambahkan model LSTM dengan metode *attention* yang diusulkan oleh *Bahdanau* (Bahdanau et al., 2015) dan *Luong* (Luong et al., 2015). Target yang ingin dicapai pada tahapan ini adalah perbaikan *caption* dengan nilai *BLEU* lebih tinggi dibandingkan model *captioning* citra tanpa *attention*. Hasil *captioning* yang diperoleh diukur kembali dengan BLEU skor untuk melihat perubahan *caption*.

Tahap Penelitian empat. Tahap ini adalah melakukan eksperimen dengan mengusulkan perubahan *backbone CNN* dan model *Transformers* sebagai *model decoder*. Model ekstraksi citra diusulkan menggunakan *Xception* sebagai *backbone* ekstraksi citra (Chollet, 2017). Selain model *backbone*, model pembangkit kata juga menggunakan *Transformers* (S. He et al., 2020; G. Li et al., 2019; L. H. Li et al., 2019). Pendekatan model *Transformers* menjadi sasaran eksperimen untuk memperoleh prediksi kata untuk *caption*. Dampak penggunaan *Transformers* pada model *captioning* citra menjadi sasaran telaah dalam model pembangkit kata citra geologi bebatuan. Hasil *caption* tetap diukur dengan menggunakan skor *BLEU* dan skor RougeL.

4. Hasil dan Pembahasan

Eksperimen yang dilakukan ini menggunakan komputer generasi dua belas dari I7. Fasilitas yang digunakan adalah *python* ver 3.8, *tensorflow ver* 2.x dengan GPU dan RAM 32GB dan bisa diperbesar sesuai aplikasi. GPU yang digunakan adalah *GeForce RTX 3050*. Library yang digunakan untuk melakukan eksperimen ini seperti *numpy*, *pandas*, *string*, *pickle*, dan *os*. Sedangkan library yang digunakan untuk membangkitkan model adalah *keras 2.3.0* dan *tensorflow 2.x*.

Tabel 2 Spesifikasi komputer

Configuration Item	Value
CPU	Processor Intel(R) Core(TM) i7-12700H CPU @ 2,3 GHz, 20 Core(s), 16 Logical Processor(s)
Graphic Processor Unit	NVDIA GeForce RTX3050, 2304 CUDA cores
Memory	DDR6 4GB
Solid Stated Disk	512 GB
Pyhton	3.8.5
Tensorflow	2.5.0

Hasil yang disajikan adalah rekayasa model *caption CNN* untuk ekstraksi fitur. Rekayasa ini untuk melihat sejauh mana *captioning* usulan bisa lebih baik dari *baseline*. Rekayasa model ini merupakan usulan kebaruan dari model *caption* yang berbeda dengan *baseline*. Seperti yang telah disampaikan pada bab pertama bahwa tujuan dari penelitian adalah mendapatkan kebaruan mengenai model *captioning* dengan menggunakan Xception (Chollet, 2017) dan *Transformer* (Vaswani et al., 2017a). Model mesin yang diusulkan ini diberikan penyebutan yaitu VaT (*Visual Attention*) untuk *Xception* dan *SeTrans* (*Semantic Transformers*) untuk *Transformers*.

Kebaruan yang diusulkan adalah merekayasa pola CNN dengan model *separable* CNN. Teknik *separable* CNN digunakan untuk mengefesienkan proses identifikasi

fitur *edge* dengan memberikan hasil jumlah parameter yang lebih sedikit dibandingkan operasi regular CNN. Keunggulan yang diharapkan dengan menggunakan *separable* CNN seperti efesiensi waktu dan jumlah parameter train hasil dari proses separable *convolution*. Pada waktu melakukan identifikasi *edge*, metode separable CNN ini dapat berlangsung lebih cepat dan mendapatkan hasil yang sama dengan regular CNN. Model rekayasa *caption* dengan menggunakan *Transformers* bisa dikerjakan dengan memperhatikan peletakan mekanisme *Transformers*. Model transformer pada awalnya digunakan untuk mesin translasi *text-to-text*. Kemudian model ini diterapkan pada *captioning* model untuk *generate caption* (J. Li et al., 2019).

Encoder merupakan bagian untuk menghasilkan peta fitur dengan menggunakan pendekatan Xception (Chollet, 2017). Bagian *decoder* adalah untuk mendapakan *caption* dengan proses transformer (Lee et al., 2020). Penelitian yang dilakukan pada klasifikasi citra batuan, diperoleh beberapa kemungkinan yang bisa ditargetkan pada penelitian mengenai CNN. Parameter CNN seperti jumlah lapis konvolusi, ukuran *filter*, ukuran *stride*, *pooling*, fungsi aktivasi, regularisasi, jumlah *fully connected layer*, dan fungsi *softmax*.

Pembuktian hipotesis pada disertasi ini adalah melihat sejauh mana signifikansi rekayasa model sehingga menghasilkan skor BLEU yang berkualitas tinggi. Rekayasa model yang dilakukan adalah berdasarkan ekstraksi fitur citra dengan struktur separable CNN, pembangkit kata *transformers*, dan citra berukuran 299x299 piksel. Pada pembangkitan *caption* digunakan *transformers* sebagai model bahasa dan masukan adalah nilai vektor dari hasil metode *word2vec word embedding* (Mikolov et al., 2013). Pemilihan *word2vec* sebagai *word embedding* didasarkan atas keberhasilan meningkatkan nilai skor BLEU pada eksperimen *Visual Attention* (VaT) (Chollet, 2017) dan *semantic transformer* (SeTrans) (Lee et al., 2020; L. H. Li et al., 2019; Weijie et al., 2020).

Eksperimen yang dilakukan melihat sejauh mana pengaruh rekayasa jumlah lapis dan parameternya terhadap *caption* pada domain yang diteliti. Captioning citra merupakan konfigurasi rekayasa dari model CNN dengan *tuning* parameter pada banyak layer CNN, *filter*, *stride*, metode *padding*, dan lapis *pooling*. Rekayasa ini bertujuan untuk mendapatkan peta fitur dengan fitur yang detil. Beberapa parameter yang dibangun untuk rekayasa CNN adalah:

- Lapis CNN pada VaT mengikuti Xception yang diusung oleh Chollet.
- Pendekatan arsitektur terbagi atas tiga bagian yaitu *Entry Flow*, *Middle Flow, dan Exit Flow*.
- *Entry flow* merupakan bagian lapis awal dengan dimulai masukan vektor berukuran 299x299 yang selanjutnya diproses dengan convolusi dengan luaran peta fitur berukuran 19x19x728.

- *Middle flow* merupakan layer lanjutan yang menerima masukan dari *entry flow* dengan vektor berukuran berukuran 19x19x728 dan memberikan *output shape* 19x19x728 dengan melakukan pengulangan sebanyak delapan kali.
- *Exit flow* adalah bagian terakhir dari *Xception* dengan melakukan proses masukan peta fitur dari *middle flow* dengan ukuran shape 19x19x728. Pada tahap akhir dilakukan *pooling* dengan menggunakan fungsi *Global Average Pooling* sehingga menghasilkan unit sejumlah 2048. *Flatten* unit berjumlah 2048 ini akan digunakan sebagai masukan untuk LSTM maupun *Transformers* sebagai pembangkit kata.
- Pendekatan fungsi aktivasi ReLU memungkinkan nilai setiap matrik tetap pada informasi yang dominan.
- Penggunaan Optimizer ADAM sebagai fungsi SGD untuk mendapatkan nilai *local* dan *global* minimum dari suatu gradien.

Jumlah parameter latih yang diperoleh pada setiap eksperimen mendapatkan sejumlah parameter latih yang berbeda untuk setiap mesin, seperti yang ditunjukan pada Tabel 1.

Tabel 1 Jumlah Paratemer Train dari setiap Model Ekstraksi fitur Gambar

Model	**Xception**	**VGG16**
Trainable parameters	23,626,728	138,357,544

Hasil yang ditunjukkan pada Tabel 2 bisa dikatakan bahwa setiap rekayasa pada CNN akan mendapatkan parameter latih. Hal ini bergantung pada pendefinisian setiap *filter* dan *channel* yang ditetapkan. Model ini juga menghasilkan skor BLEU yang berbeda (Papineni et al., 2002). Tetapi, hasil yang diberikan dapat melebihi dari *Baseline*.

Tabel 2 Hasil BLEU dari model yang diusulkan

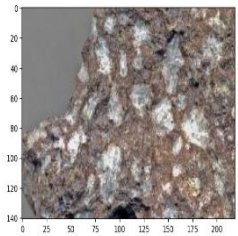
Model	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU-4**
VaT (Xception) – LSTM-word2vec	0.933	0.843	0.743	0.542
VaT (Xception) – Trans-word2vec	0.908	0.877	0.750	0.510
VaT (Xception) – GRU-word2vec	0.919	0.880	0.791	0.582

Eksperimen dalam membangkitkan *caption* menggunakan *Transformers* sebagai bagian dari *decoder* dapat dilakukan dengan hasil BLEU tertera pada Tabel 2. Tampak pada Gambar 4 bahwa BLEU yang dihasilkan masih tinggi yaitu di atas 40%. SeTrans sebagai *decoder* yang memproduksi kata memberikan panjang kata

yang lebih sedikit dari *caption* referensi. Efesiensi dalam memproduksi kalimat memberikan hasil yang menuju objek tersebut.

Prediksi:
batupasir batuan sedimen klastik dan besar lapuk
BLEU-1: 0.7143; BLEU-2: 0.5976;
BLEU-3: 0.5215; BLEU-4: 0.4347
Referensi:

1. Singkapan batuan sedimen klastik dengan bidang perlapisan yang tidak tegas, batulumpur karbonatan, masif, retak-retak, sebagian hancur dan mulai lapuk
2. Singkapan batuan sedimen klastik dengan bidang perlapisan yang tidak tegas, masif, retak-retak, sebagian hancur sehingga mulai lapuk dan batulumpur karbonatan
3. Singkapan batuan sedimen klastik dan batulumpur karbonatan
4. batulumpur karbonatan dan Singkapan batuan sedimen klastik
5. Singkapan batuan sedimen klastik dengan bidang perlapisan yang tidak tegas dan batulumpur karbonatan



Prediksi:
andesit berwarna coklat
BLEU-1: 1.00; BLEU-2: 1.00; BLEU-3: 1.00
Referensi:

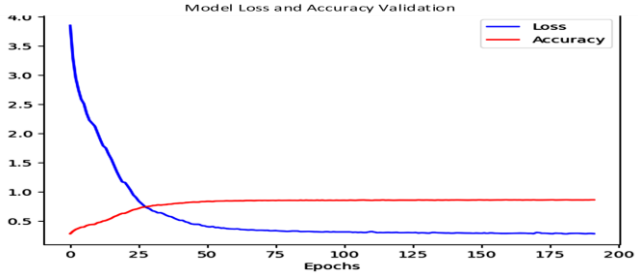
1. andesit berwarna coklat yang bertekstur porfiritik
2. andesit bertekstur porfiritik
3. andesit berwarna coklat
4. andesit coklat yang bertekstur porfiritik
5. batuan andesit coklat bertekstur porfiritik

Gambar 4 Hasil *caption* dari VaT dan *Transformers*

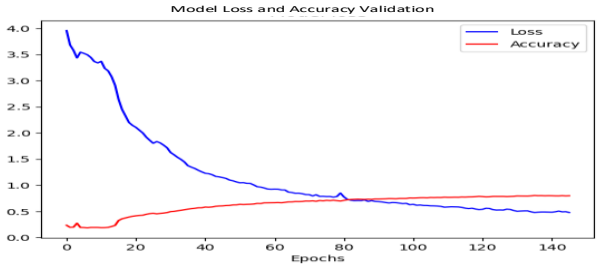
Gambar 5 dan Gambar 6 menujukkan pergerakan dari setiap perhitungan fungsi *loss* dan *accuracy* dengan fungsi *cross categorical entropy*. Pada Gambar terlihat bahwa nilai *accuracy* menaik pada sekitar epoch 25 ke atas.

Pada Gambar 6 terlihat bahwa nilai *accuracy* naik pada epoch 80 ke atas. Perbedaan ini disebabkan karena perbedaan fungsi dari masing-masing model. *Transformers* memiliki dua task yaitu *multi-head attention* dan *dot product*. Task ini mengakibatkan nilai *gradient descent* yang dihitung lambat untuk mencapai

konvergen. Kestabilan *accuracy* untuk nilai BLEU pada kedua mesin ini dicapai pada saat epoch di atas 100. Hal ini membuktikan bahwa LSTM dan *Transformers.*



Gambar 5 Grafik fungsi validation Loss dan Accuracy VaT-LSTM



Gambar 6 Grafik fungsi loss dan *accuracy* VaT-*Transformers*

Pembahasan hasil ini meliputi arsitektur CNN pada setiap model seperti VGG16, Xception, SeTrans, dan *word embedding* terhadap *image description* geologi yang dihasilkan. Pertimbangan pemilihan *baseline* model VGG16 bahwa model ini merupakan model paling dasar dari pengembangan *caption*. Beberapa model baru sering membandingkan hasil dengan VGG16 tersebut (Karpathy and Fei-Fei, 2015; Mullachery and Motwani, 2018). Pengembangan yang dilakukan adalah model arsitektur, seperti *task* pada ekstraksi fitur citra seperti ResNet50, dan InceptionV3 menjadi perhatian untuk dijadikan model rekayasa. Hal ini disebabkan kestabilan dalam menghasilkan ekstraksi. Juga kecepatan dalam mendapatkan akurasi dan mereduksi *loss* ketika melakukan ekstraksi fitur citra (Bhatia et al., 2019; Chollet, 2017; Szegedy et al., 2016).

Tabel 3 disampaikan perbandingan penggunaan teknik antara VGG16, dan model yang diusulkan dilihat dari arsitektur CNN yang disusun. Arsitektur VGG16 tidak menggunakan normalisasi agar dalam proses ekstraksi tetap mempertahankan jumlah parameter yang diproses. Hal ini dilakukan agar konsistensi akurasi bisa tetap terjaga. Tabel 3 menunjukkan bahwa Xception memiliki waktu ekstraksi lebih lama

dibandingkan arsitektur CNN(3,3). Penggunaan optimizer ADAM membantu mempercepat dalam ekstraksi fitur. Jika dilihat dari waktu pada Tabel 3, mulai proses ekstraksi sampai dengan menghasilkan *caption*, model VaT-SeTrans terkonfirmasi memiliki waktu perolehan ekstraksi lebih cepat dari VGG16.

Tabel 3 Eksperimen CNN sebagai ekstraksi citra

Model Ekstraksi fitur Gambar	**Layer**	**Fil-ter**	**Paramaters**	**Size In-put**	**learning Rate**	**Pooling**	**Waktu (menit)**
Vgg16	16	1x1, 3x3, 5x5	138,357,544	224x 224	ADAM (SGD) Optimizer, lr = 0.0001	Max Pooling	243.73
CNN(3,3)	6	3x3	4,310,144	224x 224	ADAM (SGD) Optimizer, lr = 0.0001	Max Pooling	78.97
Xception	16	3x3	23,626,728	224x 224	ADAM (SGD) Optimizer, lr = 0.0001	Max Pooling	113,75

Perhitungan *loss* yang diperoleh berdasarkan persamaan $\mathrm{CE} = -\sum_{i}^{C} t_i \log(f(s)_i)$ telah menunjukan bagaimana penggunaan teknik optimisasi *log likelihood* terhadap parameternya. *Loss* yang ditampilkan menyatakan adanya perbedaan optimasi antara *argmax* dari parameter dengan model yang dihasilkan. Perbedaan ini berdasarkan distribusi empirik yang didefinisikan dengan pelatihan set dan distribusi probabilitas dari model yang dihasilkan. Perbedaan untuk setiap *loss* pada setiap mesin tidak terlalu signifikan (Chollet, 2017; S. He et al., 2020) .

Tabel 4 kolom BLEU 1 - 4 disampaikan perhitungan hasil dari dataset validasi. Perhitungan BLEU menitikberatkan pada kepresisian dari model bahasa yang digunakan. Penggunaan *tranformers* dan *attention* sebagai model bahasa untuk pembangkitan *caption* pada dataset validasi telah mencapai nilai antara 40% - 60% untuk *word embedding* menggunakan *word2vec*. Pada model yang dikembangkan dengan *word2vec* dan *attention* menghasilkan nilai BLEU yang lebih tinggi dari *baseline model*. Proses *caption* sangat menitikberatkan pada keteraturan kata yang dibangkitkan, pencocokan hasil prediksi *caption* dengan *caption referensi* memiliki kedekatan hasil yang cukup signifikan. Hasil dari model yang dikembangkan dengan menggunakan *word emnbedding word2vec* memiliki keberhasilan *caption* yang mendekati *referensi*. Pendekatan Xception sebagai ekstraksi fitur citra terkonfirmasi mengungguli model mesin yang diusulkan. Pada model VaT dan SeTrans terkonfirmasi bahwa nilai BLEU-4 yang diperoleh melebihi *baseline model* yang digunakan. Kejadian ini memicu nilai BLEU skor menjadi kecil, diketahui hasil nilai

BLEU pada model VaT dan SeTrans adalah BLEU-1= 0.908, BLEU-2=0. 877, BLEU-3= 0. 750, dan BLEU-4=0. 510.

Berdasarkan eksperimen yang telah dilakukan dengan model *baseline* dan usulan model menggunakan dataset citra geologi bebatuan, ada beberapa pembahasan yang dapat dijadikan jawaban hipotesis untuk penelitian. Pembahasan hasil ini meliputi arsitektur CNN pada setiap model seperti VGG16, Xception, SeTrans, dan *word embedding* terhadap *image description* geologi yang dihasilkan. Pertimbangan pemilihan *baseline* model VGG16 bahwa model ini merupakan model paling dasar dari pengembangan *caption*. Beberapa model baru sering membandingkan hasil dengan VGG16 tersebut (Karpathy and Fei-Fei, 2015; Mullachery and Motwani, 2018). Pengembangan yang dilakukan adalah model arsitektur, seperti *task* pada ekstraksi fitur citra seperti ResNet50, dan InceptionV3 menjadi perhatian untuk dijadikan model rekayasa.

Tabel 4 Perbandingan BLEU-N skor

Encoder	Word Embedding	Decoder	Att.	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU-4**
Vgg16 (Baseline)	One Hot Vector	LSTM	-	0.532	0.489	0.490	0.438
Vgg16	Word2vec	LSTM	-	0.885	0.806	0.701	0.521
Vgg16	Word2vec	LSTM	Bahdanau	0.886	0.776	0.660	0.476
Vgg16	Word2vec	LSTM	Luong	0.854	0.729	0.607	0.455
Vgg16	Word2vec	Transformers	-	0.861	0.808	0.691	0.520
Resnet50	One Hot Vector	LSTM	-	0.704	0.683	0.684	0.655
InceptionV3	Word2vec	LSTM	-	0.660	0.638	0.634	0.595
Xception	Word2vec	GRU		0.919	0.880	0.791	0.582
Xception	Word2vec	LSTM	-	0.933	0.843	0.743	0.542
Xception	Word2vec	Transformers	-	0.908	0.877	0.750	0.510

Pada Tabel 3 disampaikan perbandingan penggunaan teknik antara VGG16, dan model yang diusulkan dilihat dari arsitektur CNN yang disusun. Arsitektur VGG16 tidak menggunakan normalisasi agar dalam proses ekstraksi tetap mempertahankan jumlah parameter yang diproses. Hal ini dilakukan agar konsistensi akurasi bisa tetap terjaga.

Xception sebagai ekstraksi fitur citra terkonfirmasi mengungguli model mesin yang diusulkan. Pada model VaT dan SeTrans terkonfirmasi bahwa nilai BLEU-4 yang diperoleh melebihi *baseline model* yang digunakan. Diketahui hasil nilai BLEU pada model VaT dan SeTrans adalah BLEU-1= 0.908, BLEU-2=0. 877, BLEU-3= 0. 750, dan BLEU-4=0. 510.

Tabel 5 Perbandingan BLEU dan RougeL

Model	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU-4**	**RougeL**	**Meteor**
VaT (Xception) – LSTM-word2vec	0.918	0.896	0.794	0.566	0.670	0.623
VaT (Xception) – Trans-word2vec	0.912	0.835	0.702	0.442	0.673	0.614
VaT (Xception) – GRU-word2vec	0.919	0.880	0.791	0.582	0.682	0.644

Pada Tabel 5 ditunjukan hasil BLEU dan RougeL dari arsitektur *captioning* citra yang diusulkan. BLEU dan RougeL menunjukan hasil *caption* dari sisi ketepatan dan perolehan hasil. Nilai BLEU-4 dapat dikatakan bahwa model dengan *backbone separable CNN* dapat lebih presisi dalam hal rekognisi citra. Model bahasa yang digunakan seperti GRU, LSTM, ataupun *Transformers* sama-sama memberikan dukungan hasil dengan capaian nilai BLEU yang berkualitas. Nilai BLEU dan RougeL memiliki kemiripan dalam pengukuran hasil (C.-Y. Lin, 2004; Papineni dkk., 2002). Hasil yang ditunjukkan RougeL menghasilkan kalimat berkualitas juga, yaitu di atas 50%. Hal ini dapat dikatakan bahwa *caption* yang diprediksi oleh arsitektur yang diusulkan memiliki ketepatan dan ketersediaan kata untuk *caption* yang dibangkitkan.

Model VaT-LSTM dan VaT-GRU memiliki kemiripan hasil dalam hal perubahan hitung BLEU dan Meteor. Hal ini disebabkan BLEU dan Meteor mengandalkan keteraturan susunan kalimat yang diproduksi. Pada evaluator RougeL mengandalkan kata yang terdapat pada kandidat dan referensi. Tabel 5 terlihat pada kolom RougeL, nilai VaT-Trans lebih tinggi dari VaT-LSTM. Hal ini dapat dipahami karena kandidat *caption* yang dihasilkan memiliki kata lebih banyak *overlap* dengan referensi *caption*. Evalutor RougeL lebih mengandalkan perolehan kata dibandingkan ketepatan kalimat yang diproduksi.

Ada dua hal yang perlu dipertimbangan dalam membuat model *captioning* citra secara empiris yaitu model rekognisi citra dan model bahasa. Bobot nilai yang diproduksi oleh model rokognisi citra and model bahasa akan menentukan *caption* yang paling mendekati dengan referensi. Diiketahui hasil nilai BLEU pada model VaT dan SeTrans adalah BLEU-1= 0.908, BLEU-2=0. 877, BLEU-3= 0. 750, dan

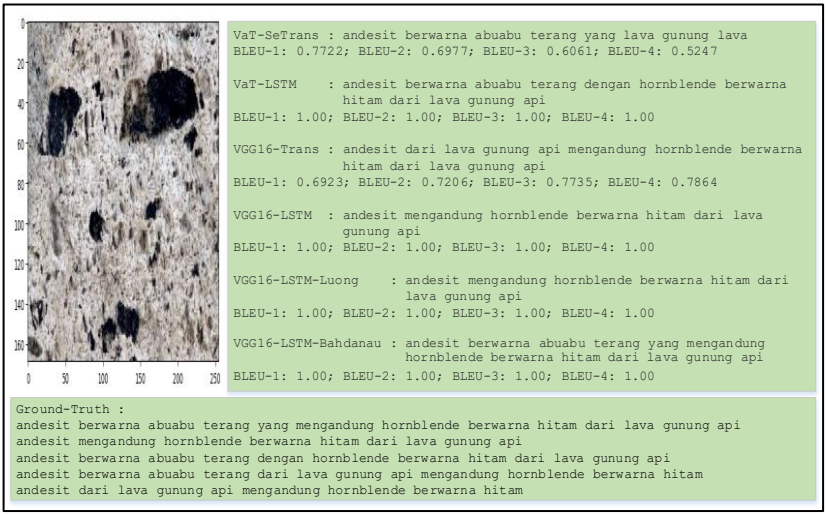
BLEU-4=0. 510. Penggunaan *Transformers* sebagai decoder untuk produksi kata dapat dikatakan kesulitan terlihat dari grafik pada Gambar (d) dan (f). Pertimbangan hyperparameter CNN dan task pada language model menjadi acuan penting untuk mendapatkan identifikasi citra latar belakang.

Pembuktian keberhasilan hipotesis dapat dilakukan dengan melakukan serangkaian eksperimen terhadap arsitektur yang diusulkan. Pada arsitektur *baseline model* dilakukan eksperimen dengan mengubah struktur *backbone* CNN. Kosakata memberi ditunjukan hasil BLEU dan RougeL dari arsitektur *captioning* citra yang diusulkan. BLEU dan RougeL menunjukan hasil *caption* dari sisi ketepatan dan perolehan hasil. Nilai BLEU-4 dapat dikatakan bahwa model dengan *backbone separable CNN* dapat lebih presisi dalam hal rekognisi citra. Model bahasa yang digunakan seperti GRU, LSTM, ataupun *Transformers* sama-sama memberikan dukungan hasil dengan capaian nilai BLEU yang berkualitas. Nilai BLEU dan RougeL memiliki kemiripan dalam pengukuran hasil (Lin, 2004; Papineni et al., 2002). RougeL memiliki hasil yang berkualitas juga, dengan hasil di atas 50%. Hal ini dapat dikatakan bahwa *caption* yang diprediksi oleh arsitektur yang diusulkan memberikan ketepatan dan ketersediaan kata untuk *caption* yang dibangkitkan.

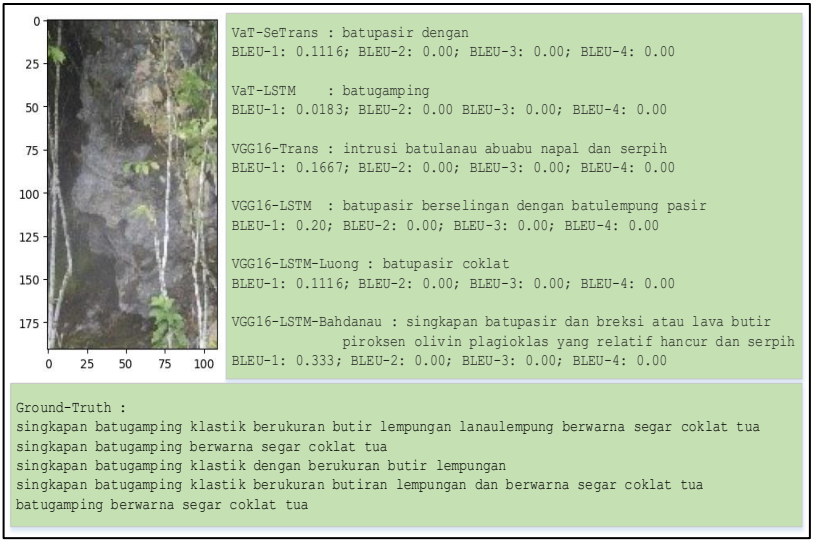
Gambar 7 menunjukkkan adanya *caption* yang yang mendekati referensi. Pada model VaT dan SeTrans sesungguhnya juga telah berhasil membuat *caption* yang mendekati referensi. Hal yang sama ditunjukkan oleh model VGG16-*Transformers*. Hasil tersebut terkonfirmasi memiliki nilai BLEU skor rendah dibandingkan model lainnya. Beberapa hal yang diperoleh ketika eksperimen dengan menggunakan VGG16 yaitu memperoleh parameter latih yang mencapai 134-jutaan fitur parameter.

Gambar 8 merupakan tampilan hasil *captioning* yang tidak tepat. Beberapa hal yang menyebabkan kesalahan ini seperti susunan layer CNN dan jumlah CNN yang digunakan belum menemukan ekstraksi yang tepat, pemilihan kata ketika proses prediksi kata dengan LSTM maupun *Transformers* belum dapat membangkitkan kata yang sesuai dengan *peta fitur*-nya. Gambar 8 ditunjukan adanya hasil validasi yang overfitting.

Pada CNN dan LSTM terlihat usulan model *captioning* untuk geologi bebatuan dengan menggunakan CNN dan LSTM terkonfirmasi masih lebih dari pada baseline dalam menghasilkan *caption*. Usulan CNN dan LSTM mengungguli usulan CNN dan Transformers sehingga *caption* yang dihasilkan memiliki BLEU skor yang tinggi. Hal demikian dapat dipahami bahwa penggunaan LSTM untuk kosakata yang jumlahnya sedikit dapat lebih efektif untuk memproduksi kata yang sesuai dengan area peta fitur. Gambar 9 dan Gambar 10 merupakan perbandingan evaluasi model berdasarkan semantic attention yang digunakan oleh model.



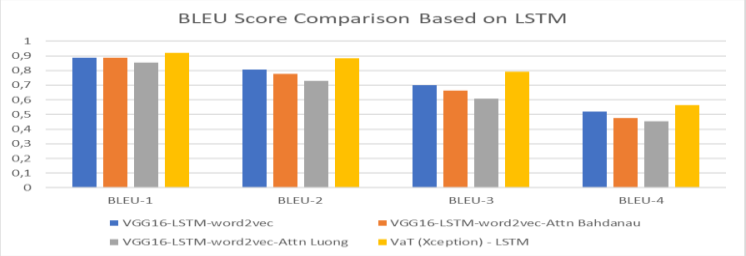
Gambar 7 Perbandingan *caption* dari berbagai Model *Captioning Citra*



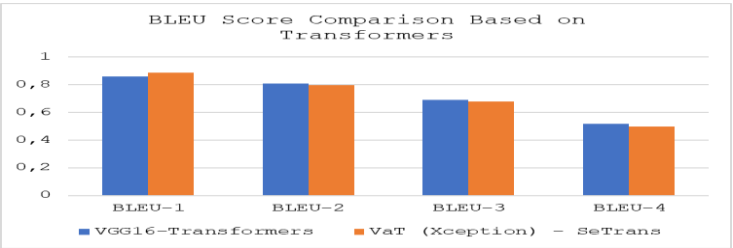
Gambar 8 *Caption overfit* yang salah ketika melakukan validasi

Setelah melakukan serangkaian eksperimen maka menjadi penting untuk membuktikan apakah hipotesis yang diajukan memiliki dampak terhadap tujuan penelitian. Dampak ini harus diuji untuk memastikan signifikansi keberpengaruhan. Pembuktian keberhasilan hipotesis dapat dilakukan dengan melakukan serangkaian

eksperimen terhadap arsitektur yang diusulkan. Pada arsitektur *baseline model* dilakukan eksperimen dengan mengubah struktur *backbone* CNN. Struktur *backbone* CNN ini akan membuktikan apakah perlakuan mengubah *hyperparameter* dan struktur jaringan dapat memberikan hasil yang lebih baik dari *baseline* arsitektur. Usulan yang diajukan yaitu struktur jaringan CNN dengan menggunakan separable CNN dan *hyperparameter* yang telah diset.



Gambar 9 Perbandingan BLEU skor pada LSTM

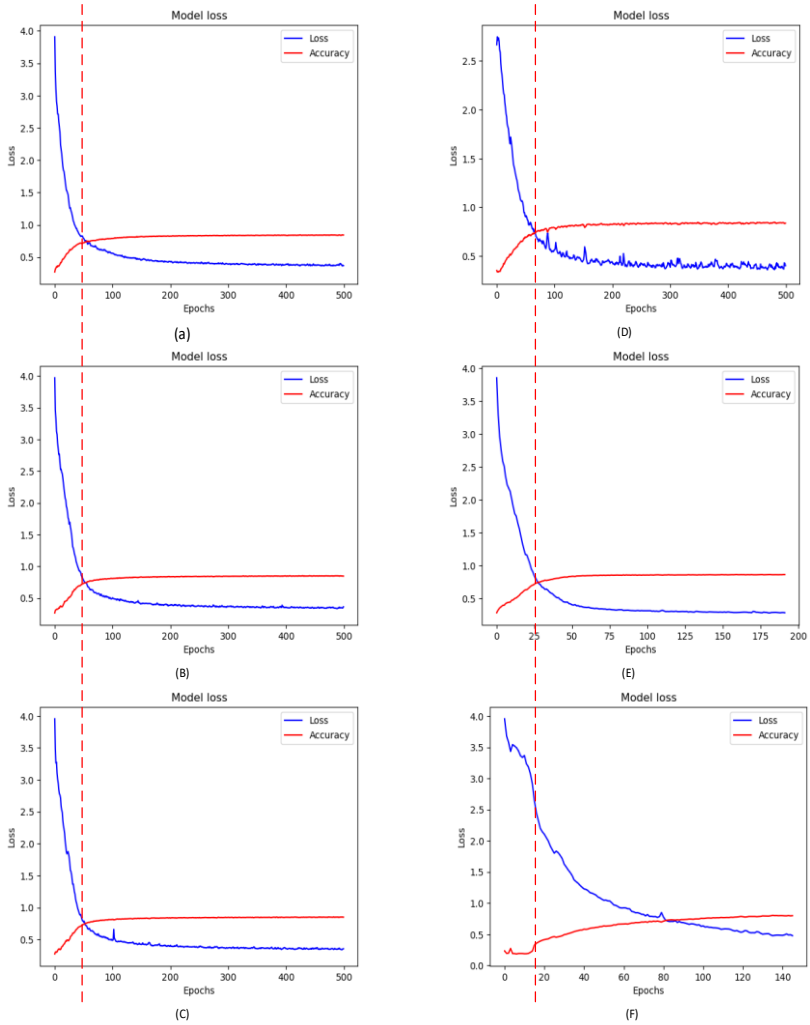


Gambar 10 Perbandingan BLEU skor pada Transformers

Pada Gambar 11 terlihat jelas bahwa penggunaan LSTM sebagai pembangkit kata dapat memberikan trend epoch terbaik ketika mencapai titik balik di bawah epoch-100. Pada penggunaan *Transformers* yaitu pada gambar Gambar 11 (d) dan (f) terlihat kesulitan ketika mencapai titik gradient descent dengan halus. Kurva menunjukan lompatan naik turun untuk mencapai akurasi terbaik ketika membentuk model *captioning* citra geologi bebatuan.

Pengujian kebenaran hipotesis yang disampaikan pada bab pendahuluan digunakan dengan pendekatan ANOVA. Analysis of variance (ANOVA) ANOVA adalah langkah awal dalam menganalisis faktor-faktor yang mempengaruhi kumpulan data yang diberikan. Analis menggunakan hasil uji ANOVA dalam uji-f untuk menghasilkan data tambahan yang selaras dengan model regresi yang diusulkan. Tes ANOVA memungkinkan perbandingan lebih dari dua kelompok pada saat yang sama untuk menentukan apakah ada hubungan diantara variabel tersebut. Hasil

rumus ANOVA, statistik F (juga disebut rasio F), memungkinkan analisis beberapa kelompok data untuk menentukan variabilitas antara sampel dan dalam sampel.



Gambar 11 Perbandingan trend mencapai epoch terbaik (a) VGG16-LSTM, (b) VGG16-LSTM-Bahdanau, (c) VGG16-LSTM-Luong, (d) VGG16-Transformers, (e) VaT-LSTM, (f) VaT-Transformers.

Tes ANOVA adalah cara untuk mengetahui apakah hasil survei atau eksperimen signifikan. Dengan kata lain, membantu penelitian untuk mencari tahu apakah perlu menolak hipotesis atau menerima hipotesis. Penggunaan ANOVA hanya untuk menyatakan bahwa hipotesis yang diajukan memang dapat dibuktikan secara saintifik dan mengandung nilai kebenaran dalam melakukan eksperimen.

Bila merujuk pada hipotesis yang diajukan bahwa perubahan arsitektur model *captioning* akan menyebabkan perubahan hasil *caption*. Pada Tabel 5 ditampilkan hasil BLEU-4 pada masing-masing model *captioning*. Sample *BLEU-4* diambil hanya untuk 145 citra gambar Alasan diambilnya nilai *BLEU-4* karena *caption* lebih terlihat dengan susunan 4-grams atau 4 kata.

Pada Gambar diperoleh hasil pengukuran hipotesis dengan menggunakan ANOVA. Hasil pengukuran dikonfirmasikan bahawa hasil P-value (P hitung) diperoleh sebesar *P-value=0.278*. Pada ketentuan uji coba statistik bila memperhatikan nilai *P-Value* yang lebih besar dari 0,05 (ukuran pembanding keberhasilan), maka BLEU-4 gabungan tidak bisa dijadikan sandaran uji statistik. Pada perhitungan lainnya untuk F-test, diketahui F-test < F-crit. Perhitungan dengan F-test untuk nilai BLEU-4 gabungan tidak bisa dijadikan uji hipotesis. Perhitungan ANOVA tersebut maka dapat dikatakan bahwa hipotesis yang menyatakan keempat nilai BLEU-4 pada setiap model tidak bisa dijadikan sandaran untuk pengujian hipotesis.

Tabel 5 Perbandingan BLEU-4 untuk 145 citra

Citra	VGG16L	VGG16B	Xcep+LSTM	Xcep+trans
1	0,0000	0,0000	0,0000	0,4347
2	0,0000	0,0000	0,0000	0,0000
3	0,0000	0,0000	1,0000	0,0000
4	1,0000	1,0000	1,0000	0,5247
5	0,0000	1,0000	0,0000	0,0000
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
141	0,0000	0,0000	0,0000	0,0000
142	0,0000	0,0000	0,0000	0,0000
143	1,0000	1,0000	1,0000	0,0000
144	0,0000	0,0000	0,0000	0,0000
145	0,0000	0,0000	0,0000	0,0000

Berdasarkan hasil uji statistik gabungan BLEU-4, kemudian dilakukan pengujian hipotesis lainnya yaitu uji statistik secara mandiri. Uji statistik mandiri artinya dilakukan pada satu model *captioning* saja, dan mengandalkan nilai BLEU-1, BLEU-2, BLEU-3, dan BLEU-4. Hasil uji statitistik *caption* dengan BLEU-4 dipilih model VaT-SeTrans yang ditunjukkan pada **Error! Reference source not found.** dilakukan pengukuran ANOVA. Pengukuran ini untuk membuktikan apakah Model VaT – SeTrans benar memiliki pengaruh dalam keberhasilan *caption*.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
VGG16L	145	66	0,455172414	0,249712644
VGG16B	145	69	0,475862069	0,251149425
Xcep+LSTM	145	81,660633	0,563176779	0,246160164
Xcep+trans	145	72,419537	0,499445083	0,240951174

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0,953187238	3	0,317729079	1,286387171	0,278166	2,620374
Within Groups	142,2681707	576	0,246993352			
Total	143,221358	579				

Gambar 12 Hasil ANOVA untuk BLEU-4

Keterangan :

Groups : merupakan nilai BLEU-4 dari masing-masing model *captioning*

Count : jumlah citra yang diujikan untuk dihitung ANOVA

SUM : jumlah seluruh nilai BLEU-4

AVERAGE : rata-rata nilai BLEU-4

Variance : nilai variasi BLEU-4

SS : sum of square (jumlah kuadrat) adalah jumlah kuadrat deviasi dari rata-rata.

Df : jumlah total observasi dikurangi jumlah batasan independen yang dikenakan pada observasi

MS : rata-rata kuadrat

F : ukuran perbedaan nilai variansi. Bila memiliki pegaruh F > 0.05

P-VALUE : ukuran pengujian pernyataan hipotesis diterima atau tidak. Nilai sandaran P-Value > 0,05.

F crit : merupakan sandaran nilai dari F Hitung (F-test), bila F test > F crit, hipotesis bisa dilakukan, jika sebaliknya maka hipotesis boleh tidak ada.

Bila mengamati Gambar 13 terdapat nilai P-Value < 0.05 yaitu $3.49x10^{-15}$. Nilai P-Value yang dihasilkan lebih kecil dari nilai pembanding yaitu 0.05. Pada Nilai F-test (F-hitung) diperoleh lebih besar dari F-crit yaitu F-test=24.69. Nilai P-value dan F-Test yang diperoleh dapat dijadikan rujukan sebagai suatu ketentuan untuk menyatakan kebenaran hipotesis. Perhitungan tersebut dapat mendukung bahwa hipotesis yang menyatakan *caption* bergantung pada aritektur model *captioning* adalah benar. Hasil ini diperoleh dengan mengandalkan Tabel 9.

Tabel 6 Nilai BLEU VaT-SeTrans

Image	BLEU-1	BLEU-2	BLEU-3	BLEU-4
1	0,7143	0,5976	0,5215	0,4347
2	1,0000	1,0000	1,0000	0,0000
3	0,2143	0,1284	0,0000	0,0000
4	0,7722	0,6977	0,6061	0,5247
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
143	0,8889	0,3333	0,0000	0,0000
144	0,0000	0,0000	0,0000	0,0000
145	0,0000	0,0000	0,0000	0,0000

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
BLEU-1	145	128,4596	0,885928	0,070581
BLEU-2	145	115,5513	0,796906	0,134975
BLEU-3	145	98,35229	0,678292	0,202601
BLEU-4	145	72,41954	0,499445	0,240951

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12,14176	3	4,047255	24,94037	3,49E-15	2,620374343
Within Groups	93,47169	576	0,162277			
Total	105,6135	579				

Gambar 13 Hasil ANOVA untuk model VaT-SeTrans

6. Kesimpulan

Terkait yang dilakukan pada penelitian ini, ada beberapa hal yang ditemukan sesuai dengan tujuan disertasi yang diajukan:

a. Penggunaan *backbone* dengan separable CNN dengan bentuk paralel arsitektur yang membedakan proses konvolusi dan perhitungan nilai residual, terkonfirmasi memberikan hasil ekstraksi citra yang dapat mengidentifikasi objek citra dengan baik. Bobot yang dihasilkan pada layer *fully connected layer* (FC) atau dense menjadi penguat dalam identifikasi objek citra. Penggunaan dense / FC layer 2048 dan 4096 menjadi output penting dalam ekstraksi citra geologi sehingga membentuk dalam menguatkan prediksi kata yang beririsan dengan fitur citra.

b. Pembangkit kata dengan menggunakan LSTM memberikan dampak signifikan menghasilkan *caption* yang terkonfirmasi lebih baik dari *Transformers*. Pembangkitkan kata dengan LSTM memberikan kalimat yang mendekati ahli geologi dengan nilai skor BLEU yang mencapai 40% dan RougeL mencapai di

atas 50%. Pada Transformers memberikan kontribusi dalam hal produksi *caption* yang lebih sederhana dibandingkan LSTM.

c. Model *Captioning Citra* yang diusulkan terkonfirmasi memiliki BLEU *skor* di atas dari *baseline model*. Uji hipotesis mengenai model, maka dapat dikatakan bahwa BLEU-1, BLEU-2, BLEU-3, dan BLEU-4 dapat dijadikan sebagai variabel uji statistik. Hasil yang diperoleh dari ANOVA merujuk pada nilai F-Test = 24.69 dan P-Value = 3.49×10^{-15}. Hipotesis yang menyatakan bahwa model arsitektur CNN dan pembangkit teks dapat memberikan *caption* yang mendekati referensi adalah benar. Pembanding pengujian ini diukur dengan F-Test > F-crit dan P-value < (P-test = 0.05).

d. Penggunaan evaluasi dengan metode BLEU skor dapat memberikan dampak usulan model lebih valid. Hal ini dibuktikan dengan evaluasi *caption* dengan menggunakan metode BLEU yang mengarahkan pada keberhasilan pembangkitan *caption*. Variabel kinerja presisi dari metode BLEU menjadi salah satu evaluasi dalam mengevaluasi peningkatan produksi *caption* yang mendekati referensi.

Kontribusi penelitian ini dapat dituliskan sebagai berikut:

1. Model rekognisi citra untuk mendapatkan peta fitur citra pada objek latar belakang dari citra bebatuan dapat memanfaatkan metode CNN. Model reguler CNN atau *separable* CNN menjadi model representasi CNN untuk ekstraksi fitur citra geologi. Citra dengan ukuran 224x224 dan 299x299 piksel dijadikan masukan untuk model rekognisi citra objek bebatuan. Ekstraksi fitur dengan pendekatan deteksi edge menjadi acuan untuk mengenali objek dalam menentukan nama bebatuan.
2. Model pembangkit kata dengan LSTM, LSTM+*Attention*, atau *Transformers* digunakan untuk membangkitkan kata yang mendekatati deskripsi citra geologi bebatuan dari ahli geologi. Panjang kata maksimal 22 kata dijadikan patokan panjang produksi kata. Metode word2vec digunakan untuk membuat *encoding* dengan domain tertutup yaitu terbatas pada kata yang ada di dataset geologi bebatuan. Pendekatan matrik dimensi 400x100 sebagai hasil dari word2vec dijadikan ukuran matrik untuk *word embedding*. Nama dan warna bebatuan dijadikan target pendekatan *semantic attention* sebagai susunan kalimat dari *caption* citra bebatuan.

6. Tindak Lanjut

Tindak lanjut penelitian ini disampaikan dalam rangka pengembangan atau keberlanjutan penelitian mengenai *captioning* citra citra geologi bebatuan. Adapun saran tersebut adalah:

1. Penentuan lapis CNN, stride, pooling, dan filter hendaknya memperhatikan susunan arsitektur CNN. Hasil unit yang dicapai dengan *separable convolutional* terkonfirmasi lebih sedikit dibandingkan proses konvolusional reguler.

2. Model pembangkit kata dengan model *long short-memory network*, masih memberikan hasil *captioning* yang *outperform* dibandingkan model *Transformers* dan *Attention* untuk kasus jumlah kata yang sedikit.
3. Model *word embedding* word2vec dengan ukuran dimensi yang sesuai dapat membantu menghasilkan prediksi *word* yang mendekati objek yang sedang diidentifikasi. Penggunaan GloVe, TF-IDF, BM25, *Sentence*-BERT, dan *Dense Passage Retrieval* (DPR).
4. *Caption* yang dihasilkan mengandalkan model kinerja presisi kata yang mendekati *referensi*, sehingga perhatian susunan hasil yang membentuk *caption* sangat dipertimbangkan.

Riwayat Hidup

Promovendus lahir di Jakarta 9 Juli 1975 dan merupakan anak ke empat dari lima bersaudara dari pasangan Solih Suhana dan Ito Suryati. Promovendus menyelesaikan sarjana tahun 1999 dari Universitas Faletehan d/h ST.INTEN Bandung. Studi magister informatika di ITB pada tahun 2003 dan lulus pada Bulan Januari 2005. Promovendus merupakan suami dari seorang istri bernama Syamsidar., S.Ag dan ayah dari tiga orang anak, yaitu Zharifatun Nur Zahra (19 Tahun), Zulfan Arif Nur Zaky (17 Tahun), dan Zimam Zuhdi Nur Muhammad (11 Tahun).

Sejak tahun 2005 – 2016, promovendus bekerja sebagai Dosen PNS LLDIKTI IV dpk Universitas Faletehan d/h ST.INTEN di Program Studi S1 Teknik Informatika. Tahun 2016 – sekarang sebagai Dosen PNS LLDIKTI IV dpk UNIKOM di Program Studi D3 Manajemen Informatika.

Beberapa projek dalam bidang IT/IS pernah ditangani :

1. Penyusunan Enterprise Arsitektur Pusat Survey Geologi (1999)
2. Rancang bangun Sistem Informasi Rumah Sakit Jiwa (2006)
3. Rancang bangun pekerjaan GIS inventarisasi jalan aspal di Kabupaten Tangerang Selatan (2010)
4. Rancang bangun inventarisasi bebatuan geologi dengan aplikasi database rock Indonesia (2000),
5. Tenaga ahli IT/IS untuk penyusunan data geologi Indonesia PSG (2019)
6. Proyek MOGEF dari Kementrian Wanita Korea (2021), (2022), (2023)
7. Proyek GeoMap dan GeoPortal Pusat Survey Geologi (2022), (2023)

Daftar Publikasi Terkait Penelitian

Nursikuwagus, A., Munir, R., dan Khodra, M. L. (2020): Image Captioning menurut Scientific Revolution Kuhn dan Popper, Jurnal Manajemen Informatika, 10 (2), 110-121, https://doi.org/10.34010/jamika.v10i2.2630

Nursikuwagus, A., Munir, R., dan Khodra, M. L. (2021): Multilayer Convolutional Parameter Tuning based Classification for Geological Igneous Rocks, International Conference on ICT for Smart Society (ICISS), 2021, IEEE. DOI: 10.1109/ICISS53185.2021.9533230

Nursikuwagus, A., Munir, R., dan Khodra, M. L. (2022): Model Caption Generator Using Visual Geometry, Residual, and Inception Architecture, International Conference on Data and Software Engineering (ICoDSE), IEEE, hal 113-118. DOI: 10.1109/ICoDSE56892.2022.9971972

Nursikuwagus, A., Munir, R., dan Khodra, M. L. (2022): Hybrid of Deep Learning and Word Embedding in Generating Captions: Image-Captioning Solution for Geological Rock Images, Journal of Imaging 2022, 8(11), 294; https://doi.org/10.3390/jimaging8110294 MDPI Publishing

Ucapan Terimakasih
Rasa terima kasih saya sampaikan kepada Dr. Rinaldi, Ir., MT, dan Dr. Masayu, ST.,MT yang telah memberikan kesempatan, nasehat, bimbingan, pencerahan, dan dorongan terus menerus hingga selesainya disertasi ini.

Ucapan terima kasih disampaikan kepada Prof. Dwi Hendratmo Widyantoro, M.Sc., Ph.D. (Alm), Prof. Dr. Carmadi Machbub, Prof. Ir. Adit Kurniawan, M.Eng., Ph.D. (Alm), Dr. Nur Ulfa Maulidevi, ST., M.Sc., Rizal Setya Perdana, S.Kom., M.Kom., Ph.D sebagai penguji pada Ujian Kualifikasi, Ujian Proposal, dan Ujian Seminar Kemanjuan I - IV dan atau reviewer buku disertasi, yang telah banyak memberikan arahan dan masukan yang sangat berharga selama proses penelitian disertasi ini dilakukan. Juga kepada Dr. Joko Wahyudiono, ST., MT, dan Fitriani Agistin, ST., M.Sc, yang telah memberikan bantuan menganotasi citra geologi.

Ucapan terimakasih juga saya sampaikan kepada Dr. Samsuri, sebagai Kepala Lembaga LLDIKTI IV Jawa Barat, Prof. Dr. Ir. H. Eddy Soeryanto.,S., MT dan Dr. Herman, Dr Wartika, dan Citra Noviyasari, MT., S.Si.,MT yang memberikan kesempatan saya untuk melanjutkan studi dan dorongan materil hingga terselesainya doktoral ini.

Penulis juga berterima kasih kepada rekan seperjuangan Program Doktor Teknik Elektro dan Informatika Angkatan 2018 serta rekan Residensi 1, Residensi 4, Lab Sistem Informasi: Dr. Zaenal, Dr. Fariska, Dr. Rini, Dr. Ilyas, Dr. Akmal, Dr. Cokorda, Jamaludin, ST., M.Sc, Dr. Heru., Dr. Edri Yunizal, S.T., M.T., Dr. Maria Irmina P., S.T., M.T., Dr. Erna Hikmawati, S.T., M.T, Kawan – kawan ITERA, UNAND, dan UNIBRAW, serta rekan-rekan lainnya, atas diskusi dan dukungannya yang sangat berharga. Saya ucapkan pula terima kasih yang sebesar-besarnya kepada seluruh pimpinan, rekan dosen, dan staf di Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia, khususnya staf dan dosen S1 Sistem Informasi serta D3 Manajemen Informartika.

Rasa terimakasih setinggi-tingginya dipersembahkan kepada Kakak, Paman, seluruh keluarga besar Solih Suhana, dan Heri Purwanto, ST.,MT.,MM.