IMPLEMENTASI MODEL *VISUAL QUESTION ANSWERING* MENGGUNAKAN *VISION TRANSFORMER* DAN *EFFICIENTNET-V2* DENGAN BERT

AHMED NIZHAN HAIKAL

2110511022

INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAKARTA

JAKARTA

2025

IMPLEMENTASI MODEL *VISUAL QUESTION ANSWERING* MENGGUNAKAN *VISION TRANSFORMER* DAN *EFFICIENTNET-V2* DENGAN BERT

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer

AHMED NIZHAN HAIKAL

2110511022

INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

JAKARTA

2025

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Ahmed Nizhan Haikal

NIM : 2110511022

Tanggal : 29 Juni 2025

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Bekasi, 29 Juni 2025

Yang Menyatakan

Ahmed Nizhan Haikal

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama	:	Ahmed Nizhan Haikal
NIM	:	2110511022
Fakultas	:	Ilmu Komputer
Program Studi	:	S-1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non - exclusive Royalty Free Right) atas skripsi saya yang berjudul:

IMPLEMENTASI MODEL *VISUAL QUESTION ANSWERING* MENGGUNAKAN *VISION TRANSFORMER* DAN *EFFICIENTNET-V2* DENGAN BERT

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Bekasi

Pada tanggal: 29 Juni 2025

Yang Menyatakan

Ahmed Nizhan Haikal

LEMBAR PENGESAHAN

Judul	: Implementasi Model *Visual Question Answering* Menggunakan *Vision Transformer* dan *EfficientNet-V2* dengan BERT
Nama	: Ahmed Nizhan Haikal
NIM	: 2110511022
Program Studi	: S1 Informatika

Disetujui oleh :

Penguji 1:

Neny Rosmawarni, S.Kom, M.Kom.

Penguji 2:

Nurul Afifah Arifuddin, S. Pd., M.T.

Pembimbing 1:

Ridwan Raafi'udin, S.Kom., M.Kom.

Pembimbing 2:

Muhammad Adrezo, S.Kom., M.Sc.

Diketahui oleh:

Koordinator Program Studi:

Dr. Widya Cholil, M.I.T

NIP. 221112080

Dekan Fakultas Ilmu Komputer:

Prof. Dr. Ir. Supriyanto, M.Sc., IPM.

NIP. 197605082003121002

Tanggal Ujian Tugas Akhir:

02 Juni 2025

IMPLEMENTASI MODEL *VISUAL QUESTION ANSWERING* MENGGUNAKAN *VISION TRANSFORMER* DAN *EFFICIENTNET-V2* DENGAN BERT

Ahmed Nizhan Haikal

ABSTRAK

Visual Question Answering (VQA) merupakan tugas *multimodal* yang memadukan visi komputer dan pemrosesan bahasa alami untuk menjawab pertanyaan berdasarkan citra. Penelitian ini mengeksplorasi pengaruh arsitektur visual (*EfficientNet V2* dan *Vision Transformer*/ViT) terhadap performa VQA, dengan BERT digunakan sebagai *backbone* teks untuk pemahaman konteks semantik *bidirectional*. Penelitian ini menggunakan *dataset* DT-VQA dan menerapkan strategi *thresholding* untuk mengatasi ketidakseimbangan label. Hasil menunjukkan bahwa model ViT-BERT dengan *fine-tuning* mencapai performa terbaik di antara varian ViT-BERT, mencatat akurasi 53,19% dan ANLS 61,95% pada *threshold* 75. Performa terbaik dicatat oleh model *EfficientNet V2*-BERT dengan *transfer learning*, dengan akurasi sebesar 56,3% dan ANLS 62,9%. Temuan ini menggarisbawahi pentingnya penyesuaian arsitektur dan strategi pelatihan untuk optimasi performa pada karakteristik data.

Kata kunci: BERT, *EfficientNet V2,* Pemrosesan *Multimodal , Visual Question Answering*, *Vision Transformer*

IMPLEMENTASI MODEL *VISUAL QUESTION ANSWERING* MENGGUNAKAN *VISION TRANSFORMER* DAN *EFFICIENTNET-V2* DENGAN BERT

Ahmed Nizhan Haikal

ABSTRACT

Visual Question Answering (VQA) is a multimodal task that integrates computer vision and natural language processing to answer image-based questions. This research explores the impact of visual architectures (EfficientNet V2 and Vision Transformer/ViT) on VQA performance, utilizing BERT as the text backbone for bidirectional semantic context understanding. We employed the DT-VQA dataset and implemented a thresholding strategy to address label imbalance. Results indicate that the ViT-BERT model with fine-tuning achieved the best performance among ViT-BERT variants, recording an accuracy of 53.19% and an ANLS of 61.95% at a threshold of 75. Overall, the highest performance was achieved by the EfficientNet V2-BERT model with transfer learning, with an accuracy of 56.3% and an ANLS of 62.9%. These findings underscore the importance of tailoring architectural choices and training strategies for optimal performance given specific data characteristics.

Keywords: BERT, *EfficientNet V2, Multimodal Processing, Visual Question Answering*, *Vision Transformer*

Kata Pengantar

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul "Implementasi Model *Visual Question Answering* Menggunakan *Vision Transformer* dan *EfficientNet-V2* Dengan BERT". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan program studi sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta. Dalam penulisan skripsi ini, penulis menyadari bahwa banyak bantuan, dukungan, dan bimbingan yang penulis terima dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua dan keluarga yang senantiasa memberikan dukungan, semangat, arahan, dan doa selama pengerjaan skripsi ini.
2. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jakarta yang telah
   memberikan izin dan kesempatan untuk menjalankan penelitian ini.
3. Ibu Dr. Widya Cholil, S.Kom., M.I.T selaku Kepala Program Studi S1-Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jakarta yang telah memberikan arahan dan dukungan selama penyusunan proposal ini.
4. Bapak Ridwan Raafi'udin, M.Kom., selaku dosen pembimbing pertama, yang telah memberikan arahan, saran, dan bimbingan selama proses penyusunan proposal ini.
5. Bapak Muhammad Adrezo, S.Kom., M.Sc. selaku dosen pembimbing kedua yang telah memberikan masukan, saran, serta dukungan selama penulisan proposal ini.
6. Teman-teman yang telah memberikan semangat dan bantuan dalam menyelesaikan skripsi.

Serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan kontribusi dalam menyelesaikan skripsi.

Daftar Isi

Daftar Gambar

Daftar Tabel

Daftar Rumus