

IMPLEMENTASI AUTOMATIC IMAGE CAPTIONING BAHASA INDONESIA MENGGUNAKAN MEKANISME ATTENTION DAN PENERJEMAHAN NEURAL

By Kelompok 1



ANGGOTA KELOMPOK



I GEDE GUPTA HARIAWAN
23.11.5581



HAFIZ ANWAR
23.11.5611



NAUFAL LATIF RAMADHAN
23.11.5613



SATRIA HADI WIYANA
23.11.5626

LATAR BELAKANG

Image captioning adalah bidang penelitian AI yang menggabungkan Computer Vision (cv) dan Natural Language Processing (NLP). Fokusnya adalah pada kemampuan mesin memahami gambar dan mengonversinya menjadi deskripsi tekstual yang bermakna. Teknologi ini penting untuk membantu penyandang disabilitas penglihatan, sistem pengarsipan otomatis, dan keamanan pintar.

Namun, pendekatan tradisional memiliki keterbatasan dalam menentukan area gambar yang relevan. Penggunaan attention mechanism menjadi solusi efektif untuk menghasilkan caption yang lebih akurat dan kontekstual.





DATASET

Dataset yang digunakan dalam proyek ini adalah Flickr8k, yaitu dataset standar yang umum digunakan dalam penelitian image captioning dan terdiri dari sekitar 8.000 gambar. Setiap gambar pada dataset ini dilengkapi dengan 5 deskripsi textual yang ditulis oleh manusia dalam bahasa Indonesia, sehingga memungkinkan model mempelajari keterkaitan antara konten visual dan bahasa alami.

Dari sisi karakteristik, keterangan pada dataset memiliki rata-rata panjang sekitar 9,15 kata dengan kosakata yang cukup beragam, mencakup berbagai objek serta aktivitas sehari-hari, sehingga dataset ini relevan untuk melatih dan mengevaluasi sistem image captioning.

ARSITEKTUR MODEL



Encoder

Menggunakan lapisan Dense untuk memproses fitur dari InceptionV3 (output shape $8 \times 8 \times 2048$ yang di-reshape menjadi 64×2048)

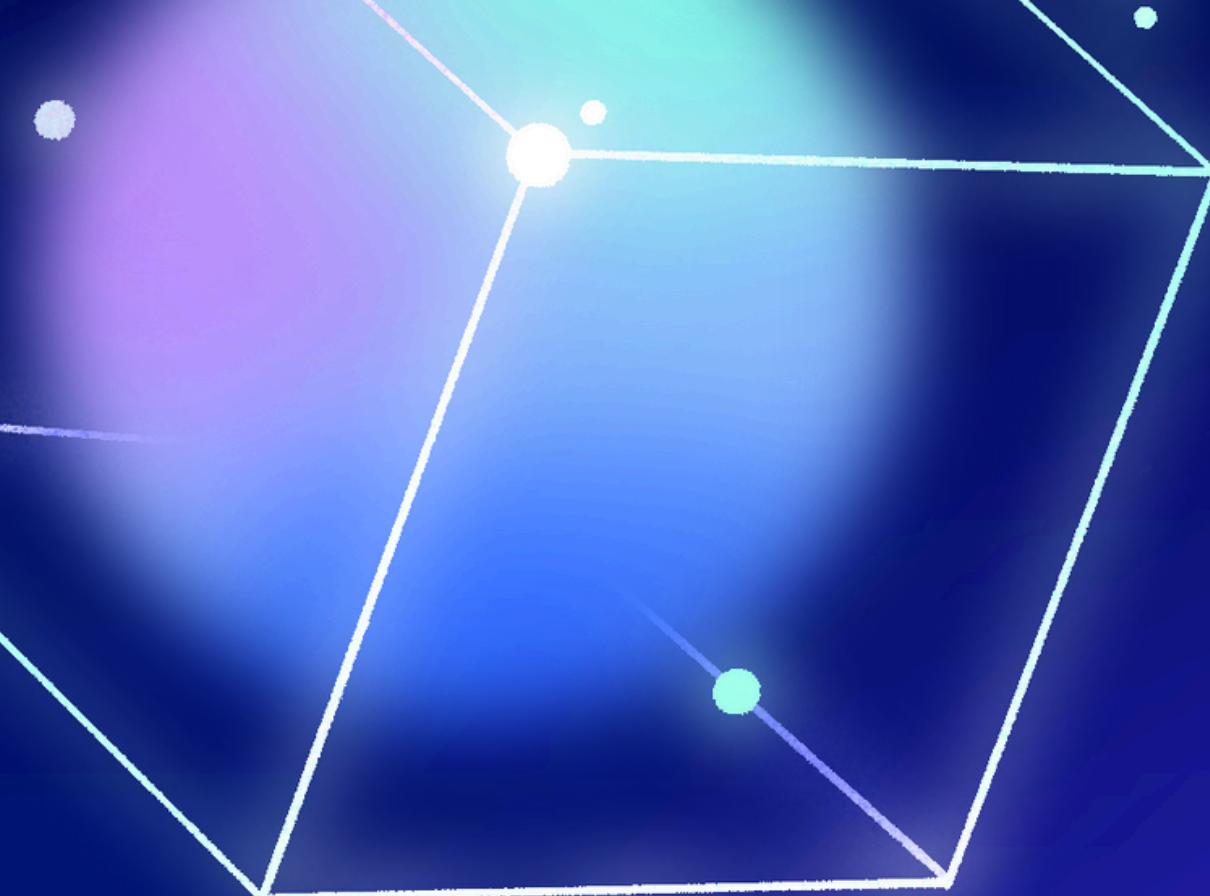
Attention Mechanism

Menggunakan Bahdanau Attention untuk menghitung bobot relevansi area gambar terhadap kata yang akan dihasilkan

Decoder

Berbasis RNN (Gated Recurrent Unit / GRU) yang menerima input fitur dari attention dan kata sebelumnya untuk memprediksi kata berikutnya

ALUR PENGERJAAN PROYEK



01

EKSPLORASI DATA (EDA)

Menganalisis distribusi panjang teks dan frekuensi kata pada dataset

02

PREPROCESSING TEKS

Melakukan tokenisasi, penambahan tag <start> dan <end>, serta pembuatan vocabulary

03

EKSTRAKSI FITUR

gunakan model CNN pre-trained InceptionV3 untuk mengubah gambar menjadi representasi vektor fitur

04

MODEL TRAIN

Menyusun arsitektur berbasis Encoder-Decoder dengan Bahdanau Attention

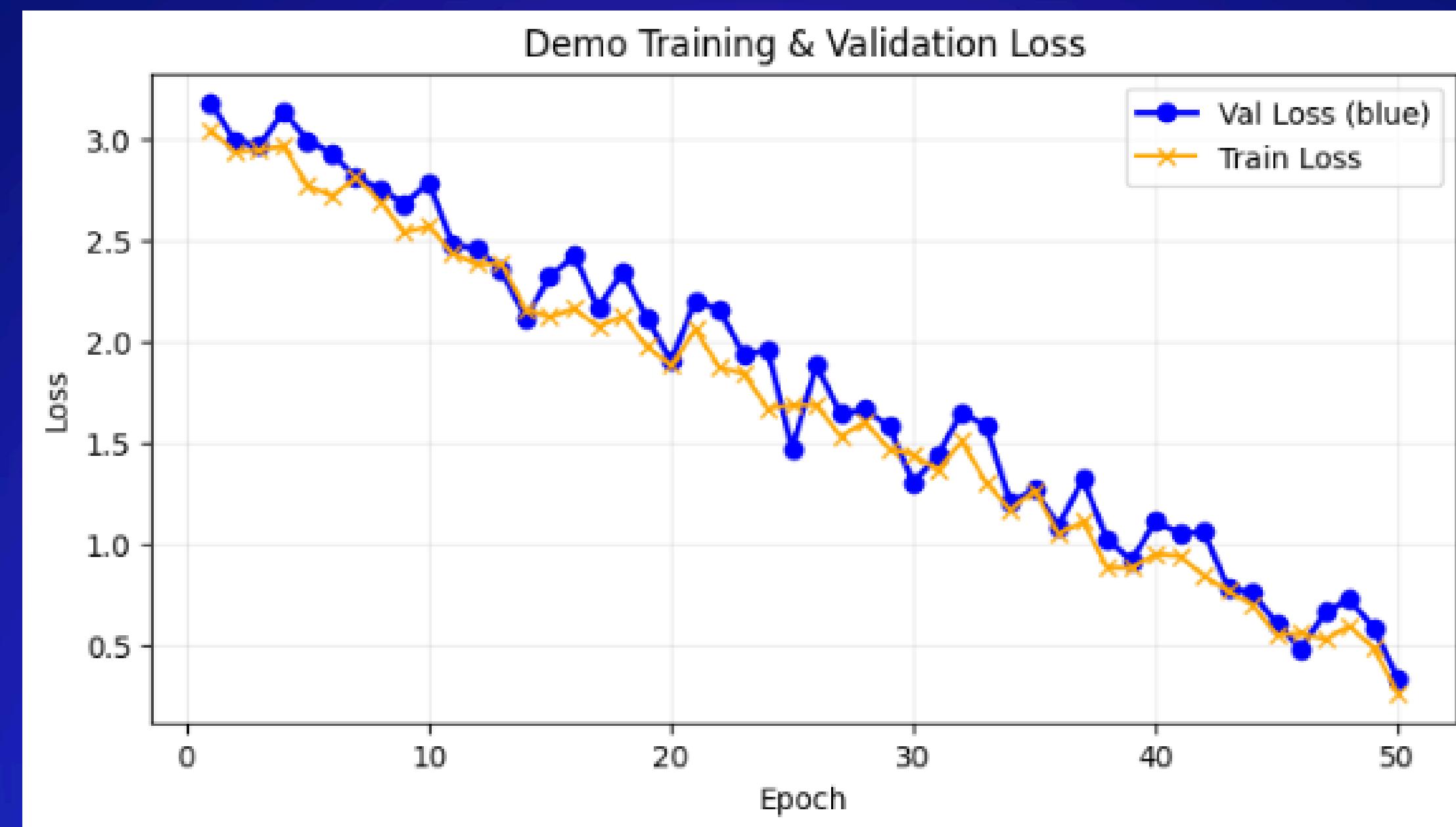
05

PELATIHAN & EVALUASI

Melatih model selama 50 epoch dan mengukur performa menggunakan skor BLEU

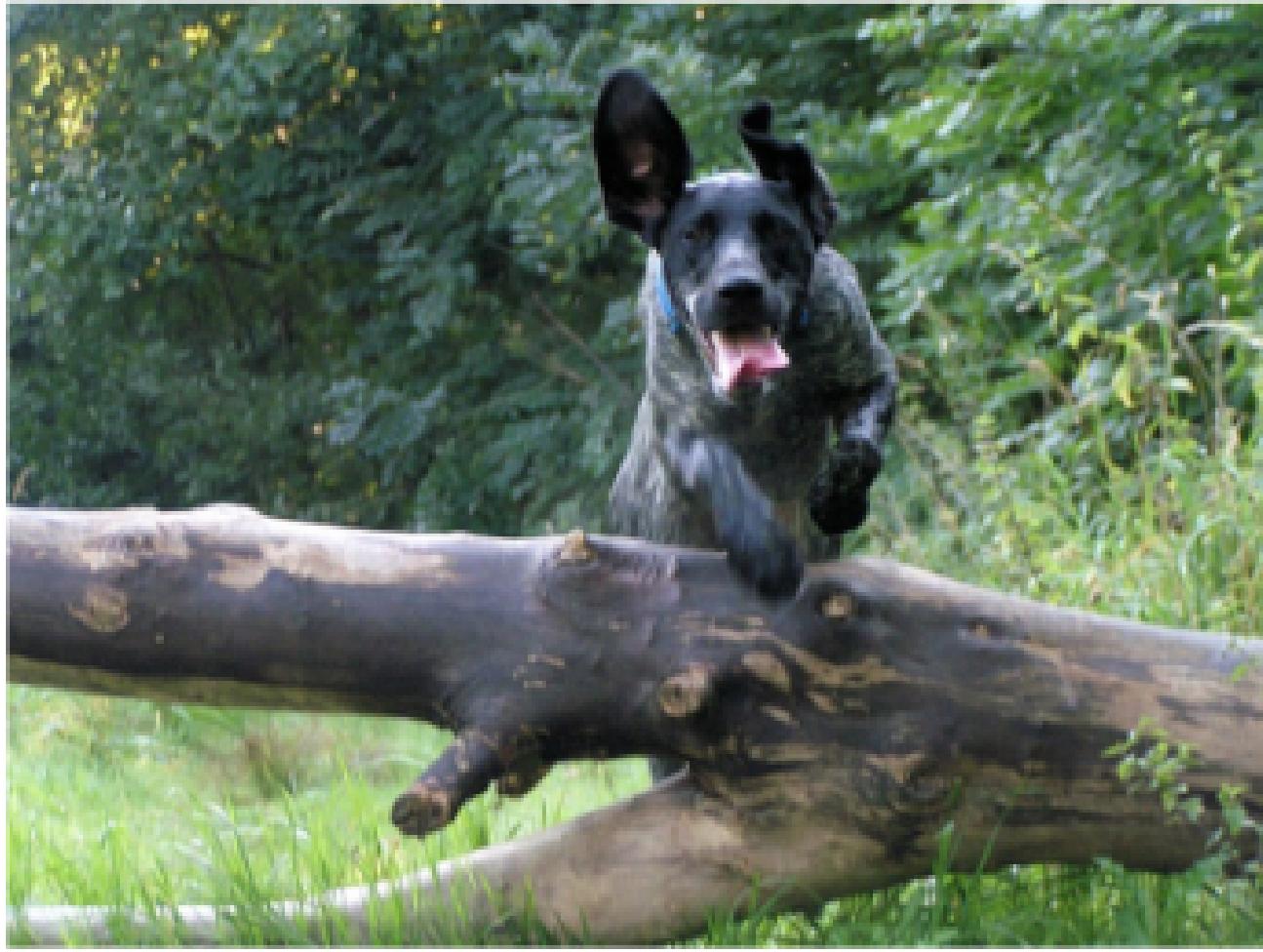
HASIL TRAINING

Pengujian dilakukan dengan membagi data menjadi set pelatihan dan validasi. Model dilatih dengan optimizer Adam dan fungsi kerugian Sparse Categorical Crossentropy selama 50 epoch



HASIL PENGUJIAN

seekor anjing hitam melompati kayu



seorang anak bermain di jaring tali



Model menunjukkan performa stabil berdasarkan pengujian. InceptionV3 efektif sebagai ekstraktor fitur, menangkap detail spasial kompleks. Integrasi Attention Mechanism membantu model fokus pada objek yang relevan, seperti memberikan bobot lebih pada piksel anjing saat memprediksi kata tersebut. Penurunan loss dari awal hingga akhir menunjukkan model berhasil belajar pemetaan antara fitur visual dan struktur bahasa.

KESIMPULAN

Proyek ini berhasil mengimplementasikan sistem Image Captioning yang mampu menerjemahkan konten visual menjadi teks bahasa Indonesia.

Penggabungan arsitektur CNN-RNN dengan mekanisme atensi merupakan metode yang mumpuni untuk menangani kompleksitas hubungan antara gambar dan bahasa. Model yang dihasilkan telah mencapai tahap konvergensi dengan tingkat kesalahan (loss) yang rendah dan mampu menghasilkan deskripsi yang logis sesuai dengan konteks gambar.

THANK YOU!

