

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Shutterstock merupakan salah satu dari banyak *agency microstock* yang melayani transaksi penjualan foto, vektor, *video* bahkan musik secara global dengan menerapkan sistem *royalty free*. Seorang yang menjual hasil karyanya pada *Shutterstock* disebut sebagai kontributor. Kontributor bisa mengunggah karyanya dalam jumlah banyak dengan berbagai macam kategori. Gambar yang di unggah akan divalidasi dan diseleksi oleh tim *Shutterstock* selama beberapa hari untuk menjaga kualitas dari karya yang akan di jual secara global. Salah satu yang menjadi pertimbangan adalah keterkaitan antara informasi deskripsi dengan gambar yang akan di unggah. Pembuatan deskripsi bisa menjadi kendala bagi kontributor yang tidak fasih dalam bahasa inggris karena deskripsi yang wajib digunakan untuk *Shutterstock* berbahasa inggris.

Kemajuan teknologi adalah sesuatu yang tidak bisa di hindari dalam kehidupan ini, karena kemajuan teknologi akan berjalan sesuai dengan kemajuan ilmu pengetahuan. Setiap inovasi diciptakan untuk memberikan manfaat positif bagi kehidupan manusia. Memberi banyak kemudahan, serta sebagai cara baru dalam menyelesaikan sebuah permasalahan biasa maupun kompleks. Perpindahan cara berkomunikasi maupun bertukar informasi di era digital membuat informasi dapat dikemas dalam bentuk gambar, text, audio, video.

Salah satu penyampaian informasi dalam bentuk visual yang dibahas dalam topik tugas akhir ini adalah gambar. Gambar memang lebih mudah untuk dipahami pada sebagian orang walaupun informasi yang disampaikan memberi pemahaman yang berbeda-beda. Teknologi membuat penyampaian informasi visual bertransformasi ke dalam bentuk digital dimana suatu file gambar memiliki deskripsi (*caption*) yang menerangkan isi dari file tersebut. Pemberian deskripsi dapat membantu pembaca memahami makna di balik gambar dan juga mempermudah mesin pencari (*search engine*) dalam menemukan gambar-gambar yang dimaksud.

Deep learning memungkinkan komputer untuk belajar secara mandiri dari kumpulan data yang diberikan (*dataset*), sehingga mesin bisa menghasilkan suatu informasi baru seperti deskripsi gambar (*generate caption*) secara otomatis.

Computer Vision merupakan model yang digunakan oleh mesin agar dapat mengenali objek yang terdapat pada gambar, sedangkan *Natural Language Processing* memberikan mesin menghasilkan rangkaian kata yang membentuk deskripsi pada sebuah gambar berupa bahasa alami dari suatu model bahasa. Salah satu teknologi *Computer Vision* yang banyak digunakan untuk pengolahan gambar adalah *Convolutional Neural Network* merupakan metode yang banyak digunakan untuk pekerjaan seperti deteksi gambar, klasifikasi gambar, dan segmentasi gambar dengan akurasi yang baik. Fitur gambar yang dipelajari oleh algoritma *Convolutional Neural Network* dapat digunakan untuk pembuatan model deksripsi gambar. Sedangkan model *Natural Language Processing* menghasilkan rangkaian deskripsi gambar dengan model *Recurrent Neural Network* dengan salah satu arsitektur yang cukup populer yaitu *Gated Recurrent Unit*. VGG19 (*Visual Geometry Group*) dengan 19 lapisan *neural network* merupakan salah satu varian dari arsitektur *Convolutional Neural Network*.

Pada penelitian *deep learning* untuk menghasilkan deskripsi dalam bahasa *Arabic* yang dilakukan oleh (Huda A. Al-muzaini, 2018) dengan menggunakan arsitektur VGG16 untuk ekstraksi fitur gambar pada dataset *Flickr* dan *Long Short Term Memory* untuk menghasilkan deskripsi dari kombinasi fitur gambar dan teks. Hasil dari BLEU score yang digunakan untuk evaluasi kinerja model didapatkan BLEU {1, 2, 3, 4} dengan nilai 46, 26, 19, 8. Hasil yang didapatkan pada BLEU score menunjukkan kinerja model cukup baik dalam menghasilkan deksripsi pada gambar.

Berdasarkan latar belakang diatas, penelitian ini menerapkan implementasi *deep learning* untuk *generate caption* gambar dengan menggunakan arsitektur VGG19 sebagai model untuk ekstraksi fitur gambar dan *Gated Recurrent Unit* untuk menghasilkan deskripsi pada gambar dari kombinasi fitur.

1.2. Perumusan Masalah

- a. Bagaimana implementasi metode *deep learning* menggunakan arsitektur VGG19 dan *Gated Recurrent Unit* dalam menghasilkan deskripsi pada gambar secara otomatis.
- b. Bagaimana tingkat akurasi yang didapatkan dari arsitektur VGG19 dan *Gated Recurrent Unit* untuk menghasilkan deskripsi gambar .

- c. Apakah *deep learning* bisa mempermudah pembuatan deskripsi gambar untuk kontributor *Shutterstock*.

1.3. Tujuan dan Manfaat

Adapun tujuan dari penelitian yang dilakukan :

- a. Membuat model untuk menghasilkan deskripsi gambar, sehingga program bisa mengidentifikasi objek di dalam gambar secara otomatis.
- b. Melakukan *preprocessing* data untuk memeriksa kualitas data gambar dan caption yang akan diuji.

Adapun manfaat yang didapatkan adalah sebagai berikut :

- a. Mengetahui tingkat akurasi yang didapatkan dari model arsitektur VGG19 dan *Gated Recurrent Unit*.
- b. Mengetahui implementasi metode *Deep Learning* menggunakan arsitektur VGG19 dan *Gated Recurrent Unit*.
- c. Mempermudah kontributor *Shutterstock* dalam pembuatan deskripsi gambar yang ingin dijual.

1.4. Ruang Lingkup

Adapun batasan-batasan pembahasan yang digunakan dalam penelitian ini adalah:

- a. Dataset yang di gunakan pada penelitian ini merupakan dataset dari situs MSCOCO 2014.
- b. Deskripsi gambar (*caption*) yang digunakan sebagai dataset merupakan deskripsi dari setiap gambar pada dataset gambar.
- c. Implementasi *generate caption* (menghasilkan deskripsi) dilakukan menggunakan *google colab*.
- d. Deskripsi yang dihasilkan berbahasa inggris.
- e. Iterasi pelatihan pada model sebanyak 20 epoch.
- f. CNN (*Convolutional Neural Network*) menggunakan arsitektur VGG19 (*Visual Geometry Group*).
- g. RNN (*Recurrent Neural Network*) menggunakan arsitektur GRU (*Gated Recurrent Unit*).

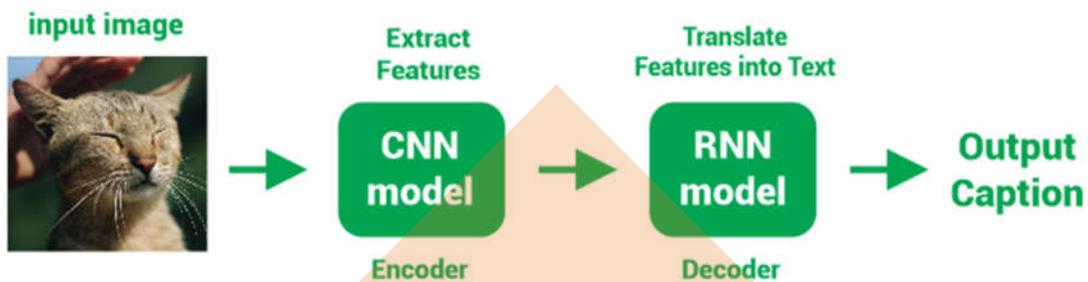
1.5. Metodologi

Metodologi yang digunakan dalam penelitian ini, terdiri dari :

a. Metode Analisis

Melakukan analisis kebutuhan sistem dimulai dari *preprocessing* pada dataset untuk mengetahui kualitas data yang akan di training dan mengidentifikasi kebutuhan informasi berdasarkan algoritma yang digunakan.

b. Metode Perancangan



Gambar 1.1. Encoder *VGG19* dan Decoder *Gated Recurrent Unit*

c. Implementasi

Program di buat menggunakan *google colab* dengan bahasa pemrograman python dan berbagai library dengan arsitektur *VGG19 (Visual Geometry Group)* varian model CNN (*Convolutional Neural Network*) yang dapat mengekstrak informasi semantik pada gambar dan *GRU (Gated Recurrent Unit)* merupakan varian model RNN (*Recurrent Neural Network*) men-*decode* fitur gambar menjadi urutan teks.

d. Pengujian

Menguji program dengan memasukkan gambar dengan berbagai kategori untuk melihat tingkat akurasi yang di dapatkan.

1.6. Sistematika Penulisan

Secara garis besar penulisan laporan Tugas Akhir ini terbagi dalam beberapa bab yang tersusun sebagai berikut :

a. BAB 1. PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan, ruang lingkup, batasan masalah, metodologi dan sistematika penulisan laporan.

b. BAB 2. LANDASAN TEORI

Uraian tentang teori yang berhubungan dengan judul tugas akhir, seperti *Deep Learning*, *Computer Vision*, *Natural Language Processing*, algoritma yang digunakan dan lain-lain.

c. BAB 3. ANALISIS DAN PERANCANGAN

Membahas mengenai analisis yang digunakan dalam membangun sistem dan merancang sistem sesuai dengan analisis yang dilakukan.

d. BAB 4. IMPLEMENTASI DAN PENGUJIAN

Memaparkan mengenai performa sistem dalam menghasilkan deskripsi gambar secara otomatis sesuai dengan gambar yang di masukkan menggunakan model VGG19 dan *Gated Recurrent Unit*.

e. BAB 5. KESIMPULAN

Pada bab ini berisi kesimpulan maupun saran dari pembahasan tugas akhir.

