

BAB V

KESIMPULAN

5.1. Kesimpulan

Penelitian ini melakukan pengujian terhadap model yang dihasilkan dari gabungan antara *CNN* dan *LSTM* dengan menggunakan *transfer learning*. Pengujian dilakukan dengan data gambar dan caption yang telah dikumpulkan melalui proses pengunduhan, baik melalui website illinois untuk data Flickr8k maupun data porn yang di ambil dari website Kaggle. Setelah data terkumpul kemudian dilakukan pemberian caption untuk data tambahan yang berasal dari Kaggle. Setelah data siap kemudian digabungkan menjadi satu dengan data Flickr8k. Data ini kemudian akan melalui proses *preprocessing* untuk menghasilkan dataset yang siap untuk di training. Adapun tahap *preprocessing* terdiri dari 3 tahap, yaitu :

1. *Bottleneck feature extraction* untuk encoder menggunakan *pretrained* model InceptionV3
2. *Onehot encoding*
3. Pembuatan data decoder

Setelah selesai *proprocessing* dan dataset sudah siap, dilanjutkan dengan mendefinisikan *hyperparameter*. Kemudian meload data encoder dan decoder. Setelah itu akan dilanjutkan dengan proses *image embedding* dan *word embedding*.

Sebelum masuk ke tahap *training* kita perlu mendefinisikan network *LSTM* berikut cell nya. Selain itu kita juga perlu mendefinisikan input output untuk setiap state dan cell nya. Kemudian mendefinisikan fungsi optimasi dan fungsi loss. Setelah pendefinisian selesai kita bisa memulai tahap *training*.

Setelah tahap *training* selesai, diperlukan pendefinisian bagaimana bentuk model yang akan disimpan. Maka akan dilakukan pendefinisian terlebih dahulu terhadap model yang akan disimpan. Baru setelah selesai kita bisa menyimpan model tersebut kedalam file keras model.

Berdasarkan grafik loss, akurasi, dan f1 bisa disimpulkan bahwa model mempunyai akurasi yang masih buruk. Hal ini disebabkan oleh data training yang masih terlalu sedikit dan nilai ekstraksi fitur yang tidak dapat dipantau sehingga weight yang didapatkan kurang optimal. Selain itu, pada penelitian ini masih menggunakan *one hot encoding* untuk ekstraksi caption nya.

Untuk proses prediksi *caption*, peneliti menggunakan metode *searching beam search*. *Beam search* sendiri memiliki parameter yang disebut *beam size* dimana seringkali dinotasikan dengan *k*. dari hasil penelitian didapatkan bahwa *beam size* dengan *k=5* dan *k=12* menghasilkan prediksi yang lebih baik. Namun, peneliti lebih menganjurkan untuk menggunakan *beam size* yang lebih besar yaitu *k=12* karena menghasilkan kandidat yang lebih banyak sehingga proses prediksi menjadi lebih baik.

Adapun proses evaluasi yang dilakukan dalam penelitian ini menghasilkan nilai *BLEU score* menggunakan metode *n-gram*. Data yang digunakan dalam proses evaluasi ini adalah data test sejumlah 1010 gambar. Dari evaluasi ini menghasilkan :

1. BLEU 1-gram = 55,84
2. BLEU 2-gram = 24,01
3. BLEU 3-gram = 10,57
4. BLEU 4-gram = 4,59
5. METEOR = 12,75
6. CIDEr = 43,84

5.2. Saran

Pada penelitian ini, masih terdapat beberapa keterbatasan dan kekurangan. Kekurangan dan keterbatasan ini bisa dijadikan acuan dan pertimbangan untuk penelitian selanjutnya. Adapun saran yang dihasilkan setelah dilakukannya penelitian ini adalah sebagai berikut:

1. Menambahkan data gambar dan caption untuk kasus data porno sehingga menambah pengetahuan fitur dan *vocabulary* daripada model yang sudah di training.

2. Menambahkan gambar kartun maupun anime sehingga semua bentuk pornografi bisa terdeteksi dengan baik.
3. Membuat caption dengan kata-kata yang lebih beragam. Hal ini bisa memperkaya kata-kata yang bisa diproses oleh *LSTM*.
4. Menggunakan *word2vector* untuk *word embedding*.
5. Penelitian ini masih menggunakan caption dalam bahasa asing. Saya berharap untuk penelitian selanjutnya ada data yang menggunakan Bahasa Indonesia.
6. Mengubah beberapa nilai *hyperparameter*. Dalam hal ini peneliti menyarankan untuk memilih nilai *learning rate* yang lebih kecil. Dan menaikkan nilai pada *batch size* dan *epochs*.
7. Menciptakan atau memodifikasi network untuk encoder sehingga proses ekstraksi fitur pada gambar porno bisa lebih optimal.
8. Penelitian ini masih menggunakan model *pretrained* sehingga agak kurang cocok untuk kasus data baru. Mengganti model *pre trained* dengan model yang di training sendiri untuk kasus data porn. Sehingga nilai *weight* dari encoder bisa dipantau.
9. Penelitian ini masih menggunakan proses *transfer learning* sehingga proses nya memakan waktu. Penggunaan konsep *RNN* secara penuh untuk proses ekstraksi fitur dan prediksi *caption* diharapkan dapat menghemat waktu training dan mempercepat proses prediksi *caption*.
10. Sistem ini dapat dikembangkan kembali untuk proses *video captioning*. Hal ini bisa digunakan untuk mengembangkan alat translasi Bahasa isyarat tangan sehingga memudahkan untuk kita berkomunikasi dengan tuna rungu dan tuna wicara.