



**PENERJEMAHAN BAHASA DAERAH BERBASIS GAMBAR
MENGUNAKAN *CONVOLUTIONAL NEURAL NETWORK***

SKRIPSI

DONNI S. SILALAH

1610511063

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2020



**PENERJEMAHAN BAHASA DAERAH BERBASIS GAMBAR
MENGUNAKAN *CONVOLUTIONAL NEURAL NETWORK***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Komputer**

DONNI S. SILALAH

1610511063

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2020

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Donni S. Silalahi

NIM : 1610511063

Tanggal : 18 Juni 2020

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

Jakarta, 18 Juni 2020

Yang Menyatakan,



(Donni S. Silalahi)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

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

Nama : Donni S. Silalahi
NIM : 1610511063
Fakultas : Ilmu Komputer
Program Studi : 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 karya ilmiah saya yang berjudul:

Penerjemahan Bahasa Daerah Berbasis Gambar Menggunakan *Convolutional Neural Network*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 18 Juni 2020

Yang Menyatakan,



(Donni S. Silalahi)

PENGESAHAN

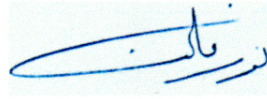
Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Donni S. Silalahi
NIM : 1610511063
Program Studi : Informatika
Judul Tugas Akhir : Penerjemahan Bahasa Daerah Berbasis Gambar
Menggunakan *Convolutionan Neural Network*

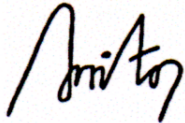
Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Henki Bayu Seta, S.Kom., MTI.
Penguji I



Noor Falih, S.Kom., MT.
Penguji II



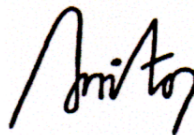
Anita Muliawati, S.Kom., MTI.
Pembimbing I



Mayanda Mega Santoni, S.Kom., M.Kom.
Pembimbing II



Informatika, M.Kom.
Dekan



Anita Muliawati, S.Kom., MTI.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 18 Juni 2020



PENERJEMAHAN BAHASA DAERAH BERBASIS GAMBAR MENGUNAKAN *CONVOLUTIONAL NEURAL NETWORK*

DONNI S. SILALAH

ABSTRAK

Indonesia dan keberagaman suku merupakan dua hal yang tidak dapat dipisahkan. Oleh karena itu, menjaga suku-suku yang ada di Indonesia wajib dilakukan oleh setiap warga nya. Salah satu upaya untuk terus menjaga sebuah suku yaitu menjamin pelestarian bahasa daerah suku tersebut. Bahasa daerah semakin jarang digunakan karena dianggap bahasa yang kuno dan masih kurangnya sarana untuk mengakses bahasa daerah tersebut. Dalam penelitian ini, akan merancang sebuah sistem untuk menerjemahkan sebuah gambar / citra berisi teks bahasa Indonesia menjadi sebuah teks dalam bahasa daerah. Penelitian ini akan melalui beberapa tahapan, dimulai dari tahap praproses, mensegmentasi karakter pada citra menggunakan pelabelan oleh *Connected Component Analysis*, kemudian tahap mengklasifikasikan karakter tersebut dengan menggunakan metode *Convolutional Neural Network*. Setelah itu, akan dilakukan penggabungan karakter menjadi sebuah teks lalu menerjemahkan nya. Pada tahap penerjemahan ini, akan menggunakan bantuan algoritma Levenshtein untuk mencocokkan teks hasil klasifikasi dengan teks dalam kamus bahasa daerah. Penelitian ini diharapkan mampu untuk menerjemahkan teks bahasa Indonesia dalam bentuk gambar menjadi teks bahasa daerah, dengan begitu akan memberi akses seseorang untuk mengetahui bahasa daerah dan mau melestarikan nya.

Kata Kunci : bahasa daerah, *connected component analysis*, *convolutional neural network*, algoritma levenshtein, penerjemahan bahasa daerah

LOCAL LANGUAGE TRANSLATION BASED ON IMAGE USING CONVOLUTIONAL NEURAL NETWORK

DONNI S. SILALAH

ABSTRACT

Indonesia and tribal diversity are two things that cannot be separated. Therefore, keeping the tribes in Indonesia must be done by each citizen. An effort to continuously maintain a tribe is to ensure the preservation of its local language. Local language is increasingly rarely used because it is considered an ancient language and still lack of means to access the local languages. In this study, would design a system to translate an image containing Indonesian language text into a local language text. This research will go through several stages, starting from the pre-process stage, segmenting the character in the image using the labeling by Connected Component Analysis, then classifying the character using the Convolutional Neural Network method. After that, all the characters will be merged into a text and then translate it. At this stage of translation, will use the help of the Levenshtein algorithm to match the text of the classification result with the text in the local language dictionaries. This research is expected to be able to translate the Indonesian text in the form of image into local language text, so it will give someone access to know the local language and want to preserve it.

Keywords : local language, connected component analysis, convolutional neural network, levenshtein algorithm, local language translation

KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul **“Penerjemahan Bahasa Daerah Berbasis Gambar Menggunakan *Convolutional Neural Network*”**.

Selama penulisan skripsi ini, penulis banyak menerima bantuan dan dukungan sehingga dapat menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Orang tua dan keluarga yang selalu memberi motivasi dan doa kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Ibu Dr. Ermatita, M.Kom., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Anita Muliawati, S.Kom.,MTI., selaku Ketua Program Studi S1 Informatika dan juga selaku pembimbing skripsi.
4. Ibu Mayanda Mega Santoni, S.Kom., M.Kom., selaku dosen pembimbing skripsi yang telah banyak memberikan ilmu, waktu dan semangat.
5. Semua teman penulis, secara khusus teman-teman Program Studi S1 Informatika Angkatan 2016 yang telah memberikan masukan dan semangat.
6. Serta semua pihak yang tidak bisa disebutkan satu persatu

Penulis berharap semoga skripsi ini dapat bermanfaat sebagaimana mestinya.

Jakarta, 18 Juni 2020

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iii
PENGESAHAN.....	iv
ABSTRAK.....	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup.....	4
1.6 Luaran yang Diharapkan.....	4
1.7 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA	6
2.1 Penelitian Terkait	6
2.2 <i>Thresholding</i>	7
2.3 <i>Connected Component Analysis (CCA)</i>	7
2.4 <i>Convolutional Neural Network (CNN)</i>	10

2.4.1	Ekstraksi Fitur (<i>Feature Learning</i>)	11
2.4.2	Klasifikasi	13
2.5	Algoritma <i>Levenshtein Distance</i>	13
BAB 3 METODOLOGI PENELITIAN		14
3.1	Kerangka Pikir	14
3.1.1	Perumusan Masalah	15
3.1.2	Studi Pustaka	15
3.1.3	Pengumpulan Data	15
3.1.4	Perancangan Sistem	16
3.1.5	Implementasi Sistem	17
3.1.6	Pengujian Sistem	19
3.1.7	Evaluasi Hasil	19
3.2	Perangkat Penelitian	19
3.2.1	Perangkat Keras (<i>Hardware</i>)	20
3.2.2	Perangkat Lunak (<i>Software</i>)	20
3.3	Jadwal Penelitian	20
BAB 4 HASIL DAN PEMBAHASAN		22
4.1	Pengumpulan Data	22
4.2	Perancangan Metode Penelitian	25
4.2.1	Praproses Data Citra	25
4.2.2	Perancangan CNN	27
4.2.3	Pembagian Data untuk Pelatihan Model CNN	30
4.2.4	Pemilihan <i>Hyperparameters</i> Model CNN	31
4.2.5	<i>Levenshtein Distance</i>	39
4.3	Implementasi Metode Penelitian	40
4.4	Pengujian Metode Penelitian	41
4.4.1	Perhitungan Akurasi	42

4.5	Analisis Hasil	45
BAB 5 PENUTUP		50
5.1	Kesimpulan	50
5.2	Saran.....	50
DAFTAR PUSTAKA.....		52
RIWAYAT HIDUP		54
LAMPIRAN		55

DAFTAR GAMBAR

Gambar 1 4-konektivitas (a) dan 8-konektivitas (b).....	8
Gambar 2 Matriks Citra Awal (a) dan Matriks Citra Baru (b)	8
Gambar 3 Pemeriksaan Awal pada Matriks Citra Awal.....	9
Gambar 4 Matriks Baru Diberi Label dan Matriks Awal Diberi 0.....	9
Gambar 5 Piksel Baru Diberi Label Berbeda	9
Gambar 6 Matriks Awal (a) dan Matriks Baru dengan 3 Objek (b).....	10
Gambar 7 Arsitektur <i>Convolutional Neural Network</i> dari LeNet-5	11
Gambar 8 Proses Konvolusi Nilai Stride 2 dan Padding 1	12
Gambar 9 Proses <i>Max Pooling</i>	12
Gambar 10 Kerangka Pikir	14
Gambar 11 Perancangan Sistem	16
Gambar 12 Proses Binerisasi pada Citra Input	17
Gambar 13 Proses Segmentasi Karakter pada Citra Biner	18
Gambar 14 Proses Pengenalan Karakter Oleh Model CNN	18
Gambar 15 Proses <i>Cropping</i> Citra Alfabet.....	23
Gambar 16 Proses <i>Cropping</i> Citra Teks	23
Gambar 17 Penamaan pada Citra Alfabet Kecil Hasil <i>Cropping</i>	24
Gambar 18 Penamaan pada Citra Alfabet Kapital Hasil <i>Cropping</i>	24
Gambar 19 Penamaan pada Citra Teks Hasil <i>Cropping</i>	24
Gambar 20 Mengubah Posisi Objek Menjadi di Tengah.....	25
Gambar 21 Data Citra Dalam Format .csv	26
Gambar 22 Praproses Citra Awal ke Biner Lalu Dilakukan CCA	26
Gambar 23 Hasil Segmentasi Objek Sesuai Label	26
Gambar 24 Penempatan Objek ke Citra Baru Ukuran 32x32 Piksel.....	27
Gambar 25 Proses Pelatihan dan Pengujian pada CNN	28
Gambar 26 Arsitektur CNN.....	29
Gambar 27 Model <i>Summary</i>	29
Gambar 28 Proses <i>K-Fold Cross Validation</i>	31
Gambar 29 <i>Classification Report</i> Hasil Uji Coba Model ke Data <i>Test</i>	38
Gambar 30 <i>Confusion Matrix</i> Hasil Uji Coba Model ke Data <i>Test</i>	39

Gambar 31 <i>Flowchart</i> Algoritma <i>Levenshtein Distance</i>	40
Gambar 32 <i>Flowchart</i> Pengujian Sistem	41
Gambar 33 Grafik Peningkatan Akurasi Klasifikasi setelah Diterjemahkan	43
Gambar 34 Grafik Peningkatan Akurasi Klasifikasi setelah Diterjemahkan	44
Gambar 35 Hasil dari Proses Segmentasi, Klasifikasi dan Penerjemahan	44
Gambar 36 Contoh Hasil Segmentasi Bagus	45
Gambar 37 Citra yang Rusak	46
Gambar 38 Kesalahan Segmentasi Akibat Citra Rusak	46
Gambar 39 Kesalahan Pelabelan Objek bukan Huruf	46

DAFTAR TABEL

Tabel 1 Operasi pada <i>Levenshtein Distance</i>	13
Tabel 2 Jadwal Penelitian	21
Tabel 3 Percobaan <i>Batch Size</i>	32
Tabel 4 Percobaan <i>Epoch</i>	33
Tabel 5 Percobaan <i>Learning Rate</i>	35
Tabel 6 Percobaan Model Terpilih Berdasarkan <i>Batch Size</i>	36
Tabel 7 Percobaan Model Terpilih Berdasarkan <i>Epoch</i>	37
Tabel 8 Percobaan Model Terpilih Berdasarkan <i>Learning Rate</i>	37
Tabel 9 Akurasi Keberhasilan Klasifikasi Berdasarkan <i>Device</i> dan <i>Font</i>	42
Tabel 10 Akurasi Keberhasilan Penerjemahan Berdasarkan <i>Device</i> dan <i>Font</i>	43
Tabel 11 Segmentasi Benar, Klasifikasi Benar dan Terjemahan Benar	47
Tabel 12 Segmentasi Salah, Klasifikasi Salah tetapi Terjemahan Benar	47
Tabel 13 Segmentasi Benar, Klasifikasi Salah tetapi Terjemahan Benar.....	48
Tabel 14 Segmentasi Benar, Klasifikasi Salah dan Terjemahan Salah	48
Tabel 15 Segmentasi Salah, Klasifikasi Salah dan Terjemahan Salah	49