Jurnal Vocational Teknik Elektronika dan Informatika
http://ejournal.unp.ac.id/index.php/voteknika/index
Vol. 11, No. 4, Desember 2024 P- ISSN: 2302-3295, E-ISSN : 2716-3989

Penerapan *Optical Character Recognition* (OCR) Dengan *Text-To-Speech* (TTS) dalam Konversi Gambar ke Suara

Prabowo Budi Utomo[1*], Ibnu Mas'ud Luthfi[2], M. Nur Fu'ad[3], M. Mujiono[4]
[1,3,4]Akademi Komunitas Negeri Putra Sang Fajar Blitar, Indonesia
[2] STAI At-Tahdzib Jombang, Indonesia
**Corresponding author* e-mail : prabowo86@akb.ac.id

ABSTRAK

Aksesibilitas informasi menjadi perhatian utama untuk memastikan bahwa semua individu dapat mengakses dan memahami konten secara maksimal Gangguan penglihatan menjadi salah satu disabilitas atau kekurangan yang cukup banyak dialami oleh orang Indonesia yang dalam perkembangannya menimbulkan berbagai masalah sebagai akibat dari kekurangan yang dimiliki salah satunya adalah aksebilitas informasi. Penelitian ini secara tidak langsung output yang dihasilkan merupakan hasil pengabungan dari menggunakan *Optical Character Recognition* dengan konversi representasi *Vector Quantized Variational Autoencoder* dengan pengubah suara *Text-to-Speech* dari *google* (gTTS) yang dilakukan sebagai upaya untuk menghasilkan kualitas suara yang lebih baik dan alami serta mempertahankan informasi asli. Hasil pengujian dalam penelitian diperoleh akurasi konversi dan pengubahan sebanyak 83,33% dengan 10 data uji dapat dikonversi dan diubah dengan baik dan cukup efektif dalam mempertahankan informasi asli dan menghasilkan suara natural.

Kata kunci *:* Akses Informasi; Gangguan Penglihatan; OCR; VQ-VAE; gTTS; Machine Learning

ABSTRACT

Accessibility to information is a major concern to ensure that all individuals can access and understand content to the fullest. Impaired vision is one of the disabilities or deficiencies experienced by quite a lot of Indonesians, which in its development creates various problems as a result of the deficiencies they have, one of which is information accessibility. This research indirectly produces the output that is the result of a combination of using Optical Character Recognition with the conversion of the Vector Quantized Variational Autoencoder representation with the Text-to-Speech voice modifier from Google (gTTS) which is carried out as an effort to produce better and more natural voice quality and retain original information. The test results in this study obtained an accuracy of conversion and conversion of 83.33% with 10 test data that can be converted and changed properly and are quite effective in retaining original information and producing natural sound.

Keywords: *Information* Access*; Visual Impairment; OCR; VQ-VAE; gTTS; Machine Learning*

I. PENDAHULUAN

Aksesibilitas informasi menjadi perhatian utama untuk memastikan bahwa semua individu dapat mengakses dan memahami konten secara maksimal. Dalam Pasal 28F Undang-undang Dasar Negara Republik Indonesia Tahun 1945, dijelaskan bahwa hak setiap orang untuk berkomunikasi dan memperoleh informasi sangat penting dalam pengembangan pribadi dan lingkungan sosial. Selain itu, setiap orang juga berhak untuk mencari, memperoleh, memiliki, dan menyimpan informasi melalui berbagai jenis saluran yang tersedia. [20], sehingga secara jelas dapat disimpulkan bahwa setiap orang tanpa memandang kondisinya dijamin oleh negara berhak mencari, memperoleh, memiliki dan menyimpan informasi.

Gangguan penglihatan menjadi salah satu disabilitas atau kekurangan yang cukup banyak dialami oleh orang Indonesia, menurut WHO pada tahun 2018 gangguan penglihatan diklasifikasikan dalam 3 kriteria [21], yaitu:

1. Deskripsi ringan mengacu pada kondisi penglihatan yang tidak terganggu oleh

gangguan tajam atau penglihatan kabur, dengan visus minimal 6/18 (logmar minimal 0,3).
2. Kondisi sedang merujuk pada gangguan tajam penglihatan dengan visus antara kurang dari 6/18 hingga 3/60 (logmar antara kurang dari 0,3 hingga 0,05).
3. Kondisi berat atau kebutaan terjadi ketika terdapat gangguan tajam penglihatan dengan visus kurang dari 3/60 hingga tidak ada persepsi cahaya (logmar kurang dari 0,05 hingga tidak ada persepsi cahaya).

Di Indonesia, terdapat lebih dari 8 juta orang yang mengalami gangguan penglihatan, dengan 1,6 juta di antaranya menderita kebutaan atau gangguan penglihatan yang parah [22], yang dalam perkembangannya menimbulkan berbagai masalah sebagai akibat dari kekurangan yang dimiliki salah satunya adalah aksebilitas informasi, namun apabila bersabar Syaikh Ibnu Utsaimin rahimahullah menasehatkan, “Mata itu adalah anggota tubuh yang amat dicintai, jika Alloh mengambilnya dan seorang itu mau bersabar dan mengharap ganjaran, maka ia akan mendapat ganti surga.”[23]. Dalam upaya meningkatkan aksesibilitas informasi, setiap individu berhak memiliki dan mengakses informasi sesuai dengan undang-undang yang berlaku. Oleh karena itu, diperlukan teknologi yang dapat berfungsi sebagai jembatan dalam mengakses informasi, terutama bagi penderita kekurangan penglihatan. Teknologi ini akan memainkan peran penting dalam memfasilitasi akses informasi bagi semua individu.

Dalam upaya peningkatan aksebilitas informasi, beberapa penelitian telah dilakukan seperti yang dilakukan oleh Wei-Ning Hsu, dkk dengan mempresentasikan model yang mampu menghasilkan gambar teks yang berasal dari pengucapan yang fasih, dimana model yang diperoleh dari pengolahan dataset MSCOCO yang diproses per unit segmen menggunakan *deep learning* (1). Dalam penelitian lainnya Isuri Anuradha, dkk memanfaatkan metode *Optical Character Recognition* (OCR) dalam pengenalan huruf yang terdapat pada berbagai tipe buku Sinhala yang diperoleh dari berbagai kuil Hindu dan berbentuk file jpg/jpeg/png/pdf, dengan dikembangkan menggunakan metode *deep learning* mampu dihasilkan model yang dapat mengenali font Malithi Web, LKLug dan model font gabungan menggunakan Noto Sans, LKLug dan Malithi Web dengan akurasi sebesar 87,07%, 87,15% dan 87,52% pada buku sinhala tipe lama (2) , maka secara tidak langsung diketahui bahwa *Optical Character Recognition* (OCR) yang dikembangkan menggunakan metode *deep learning* mampu dan cukup optimal dalam mengenali font pada dokumen berbentuk file *jpg/jpeg/png/pdf*. Dalam penelitian lain, Johanes Effendi, dkk menggunakan pendekatan yang berbeda dalam memperoleh model untuk mendapatkan informasi dalam bentuk suara, dengan mengkombinasikan jalur *Vector Quantized Variational Autoencoder* (VQ-VAE) dan MEL-SPECTROGRAM dalam membuat model *image2speech* tanpa teks sebagai upaya untuk menyediakan teknologi untuk bahasa yang tidak dikenal yang tidak ditranskripsikan (3). Dalam pemanfaatan metode *Optical Character Recognition* (OCR) Ahmed Talat Sahlol, dkk menemukan terdapat 4 fase dalam pengenalan karakter yaitu pra-pemrosesan, ekstraksi fitur, pemilihan fitur dan klasifikasi, dimana pemilihan fitur menjadi titik kunci untuk membangun sistem pengenalan karakter yang memadai (4). Hal yang berbeda dilakukan oleh Ramazan Gokay, dkk, yang memanfaatkan efek augmentasi data dan sintesis ucapan pada pengenalan suara hasil pengenalan karakter menggunakan kombinasi metode *Google Translate Text to Speech* (gTTS) dan *Deep Convolutional TTS* (DCTTS) sebagai upaya dalam menurunkan nilai *Word Error Ratio* untuk pengenalan ucapan Bahasa Türkiye yang diakibatkan dari kurangnya data *trainning* (5) .

Berdasar penelitian yang selama ini telah dilakukan serta mengacu pada permasalahan yang ada, maka dalam penelitian ini diusulkan penerapan teknologi konversi gambar ke suara memanfaatkan metode *Optical Character Recognition* (OCR) yang dikombinasikan dengan metode *Vector Quantized Variational Autoencoder* (VQ-VAE) dan dikonversi menggunakan *Google Text To Speech* (gTTS). Metode OCR diperggunakan untuk mengenali dan mengekstraksi teks dari gambar, dan selanjutnya dengan VQ-VAE untuk menghasilkan representasi diskrit dari teks hasil ekstraksi dan gTTS untuk mengubahnya representasi diskrit menjadi suara. Dengan fungsi dan output yang diperoleh dari masing-masing metode diharapkan mampu mengukur kemampuan OCR dalam mengenali dan mengekstrak teks dari gambar, serta mengukur kualitas suara yang dihasilkan dari konversi teks ke suara menggunakan gTTS dengan harapan mampu memberikan pemahaman yang lebih baik dalam penerapan OCR dan gTTS untuk konversi gambar ke suara.

Hasil penelitian ini diharapkan mampu mendapatkan informasi dari gambar dalam bentuk suara yang lebih alami dan berkualitas tinggi, sehingga mampu membantu meningkatkan aksebilitas informasi khususnya bagi penderita gangguan penglihatan.

P-ISSN: 2302-3295

II. METODE

A. Gangguan Penglihatan

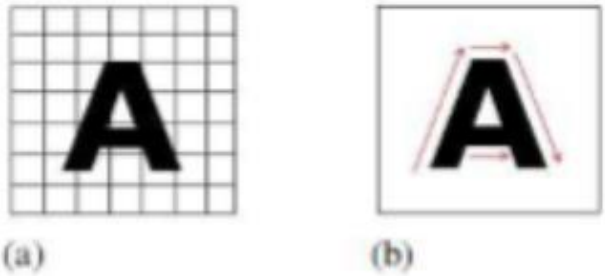
Banyak orang Indonesia yang mengalami disabilitas gangguan penglihatan, dimana sekita 8 juta orang mengalaminya dengan 1,6 juta diantaranya mengalami kebutaan [24]. Berbagai sebab dapat menyebabkan terjadinya gangguan penglihatan, seperti cedera atau penyakit mata. Gangguan penglihatan berpengaruh terhadap *visus* atau ketajaman penglihatan khususnya untuk mengetahui dan membedakan detail-detail terhadap obyek maupun permukaan (6).

Gangguan refraksi yang tidak terkoreksi yang disertai dengan katarak atau *glaucoma* menjadi jenis gangguan penglihatan yang sering ditemui (7). Gangguan ini menyebabkan mata tidak mampu fokus/melihat dengan jelas pada suatu obyek atau permukaan serta membuat pandangan menjadi kabur, dikala sudah parah akan menjadikan *visual impraiment* (melemahnya penglihatan) seperti myopia (rabun jauh), hypermetropia (rabun dekat) dan astigmatisme. Secara umum, ada beberapa faktor yang dapat menyebabkan gangguan penglihatan, katarak, yang merupakan penyebab terbanyak dengan persentase sebesar 34,47%. Selanjutnya, gangguan refraksi yang tidak terkoreksi juga menjadi penyebab kebutaan dengan persentase sebesar 20,26%. Terakhir, glaukoma juga dapat menyebabkan kebutaan dengan persentase sebesar 8,30% [25].

Lebih dari 75% kasus gangguan penglihatan dapat dihindari (6), Untuk mencegahnya, pemerintah telah meluncurkan program penanggulangan kebutaan akibat katarak di Provinsi Nusa Tenggara Barat pada tahun 2020 (7). Dalam upaya pencegahan dan deteksi dini gangguan penglihatan, *International Agency for the Prevention of Blindness* mencanangkan Hari Penglihatan Sedunia yang diperingati pada setiap bulan Oktober yang diperingati melalui beberapa kegiatan seperti pemeriksaan kesehatan mata, sosialisasi pembatasan penggunaan gawai, dan sebagainya (8) .

B. *Optical Character Recognition* (OCR)

Optical Character Recognition (OCR) merupakan Teknik yang digunakan untuk menafsirkan dokumen dalam bentuk file citra atau gambar yang dipindai menjadi teks yang dapat dibaca dna disunting oleh aplikasi komputer (9), dengan kemampuannya OCR mampu mendeteksi dan mengenali teks cetak maupun tulisan tangan dari dokumen dan mengubahnya menjadi teks yang dapat diedit (10). Terdapat dua macam pengenalan karakter didalam OCR yaitu *offline* dan *online character recognition*, dimana *offline* akan bekerja dengan mengenerate dokuken kemudian didigitalisai dan disimpan didalam komputer sebelum diproses, sedangkan *online* akan langsung memproses selama dalam proses *capture* atau pindai dokumen.



Gambar 1: *offline* dan *online character recognition* (11)

Dalam pengoperasiannya, OCR memiliki beberapa tahapan yang harus dikerjakan (12), yaitu :

2.1. Pra-pemrosesan

Tahapan pra-pemrosesan bertujuan untuk mendapatkan karakter tunggal dari teks yang dipindai dalam kondisi yang baik dan bersih, sehingga memudahkan proses pengenalan. Pra-pemrosesan dilakukan dengan melakukan normalisasi kondisi teks terlebih dahulu, yaitu dengan menghilangkan gangguan seperti titik, memperbaiki orientasi citra teks, mengubah menjadi biner, dan membagi teks menjadi segmen-segmen. Tahapan ini memiliki pengaruh yang signifikan terhadap tingkat akurasi pengenalan karakter. (12).

2.2. Ekstraksi Fitur

Tahapan ekstraksi fitur dilakukan untuk menemukan atribut pola karakter yang terpenting serta berbeda dari karakter lain supaya bias diklasifikasi [26].

2.3. Pengenalan

Pola karakter yang telah terpetakan ke dalam nilai vector selanjutnya dikelompokkan sesuai nilai yang sama atau hamper sama. Proses klasifikasi nilai vector menjadi kunci proses pengenalan karakter(13).

2.4. Paska-proses

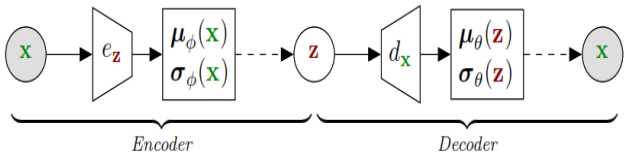
Tahap paska-proses dilakukan sebagai proses koreksi, disebabkan model yang digunakan untuk mengkoreksi kesalahan yang terjadi pada susunan kata.

C. *Vector Quantized Variational Autoencoder* (VQ-VAE)

Vector Quantized Variational Autoencoder (VQ-VAE) merupakah salah satu variasi dari *generatif Variational Autoencode*r (VAE) yang mampu menghasilkan representasi diskrit dari data input. Tujuan dari VQ-VAE adalah mengabungkan representasi laten diskrit dengan model *generatif Variational Autoencoder* (VAE) (3) .

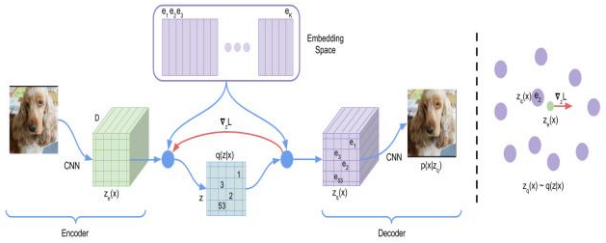
Variational Autoencoder (VAE) sebagai dasar memiliki dua komponen utama yaitu *encoder* dan *decoder*, dimana *encoder* diperggunakan dalam mengubah data input menjadi representasi laten kontinu melalui pemodelan distribusi probabilitas laten. Sedangkan *decoder* mengambil representasi

laten yang telah dibuat sebagai input dan melakukan rekonstruksi sesuai data input asli.



Gambar 2. Skema Pengolahan VAE (14)

Penggunaan konsep *Vector Quantization* ditujukan untuk memperoleh representasi diskrit sehingga mampu mempertahankan informasi penting dari data input. Konsep *Vector Quantization* melakukan penggabungan representasi laten kontinu menjadi representasi diskrit dengan mendistribusikan setiap representasi laten pada kode diskrit sesuai codebook VQ-VAE. Proses distribusi ini disertai dengan pengabungan representasi laten dengan kode terdekat pada codebook, dalam hal ini kode terdekat berisi sejumlah vektor diskrit yang memiliki jarak eculidean terdekat dengan representasi laten.



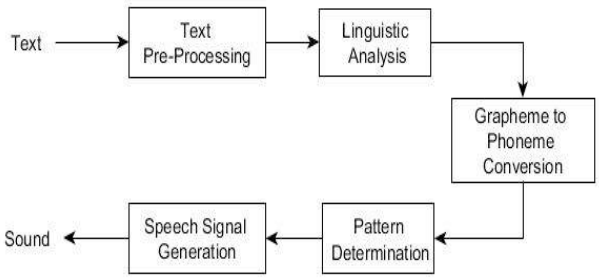
Gambar 3. Proses pengolahan VQ-VAE (15)

Keuntungan penggunaan VQ-VAE adalah kemampuan meningkatkan kecepatan komputasi dan efisiensi penyimpanan sebagai bentuk output representasi diskrit. Disamping itu, representasi diskrit memungkinkan interprestasi yang lebih mudah dan sederhana dalam melakukan klasifikasi atau pengambilan keputusan.

Dalam konteks konversi gambar ke suara, VQ-VAE dipergunakan dalam mekanisme *encoder-decoder* teks hasil ekstraksi OCR untuk menghasilkan representasi diskrit berupa potongan-potongan teks terpisah. Representasi diskrit sebagai hasil proses *encoder-decoder* diperggunakan sebagai data input model *Google Text-to-Speech* yang akan diubah menjadi suara.

D. *Google Text To Speech* (gTTS)

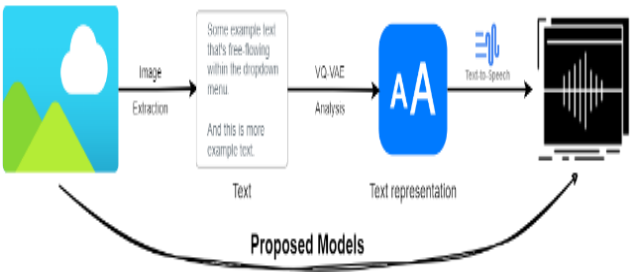
gTTS (*Google Text-to-Speech*) merupakan pustaka python yang diperggunakan untuk mengkonversi teks menjadi ucapan atau suara [27] yang dapat disimpan sebagai file mp3 (17).



Gambar 4. gTTS konversi teks ke suara (18)

Pada dasarnya gTTS menggunakan *WaveNet* untuk menghasilkan suara, arsitektur inti dari *WaveNet* merupakan model API yang mampu menghasilkan audio mentah sesuai teks yang dimasukkan. Model ini disamping menghasilkan audio mentah juga mampu mensitesis ucapan, sehingga audi yang dihasilkan mirip dengan suara manusia (19).

E. Rancangan Model



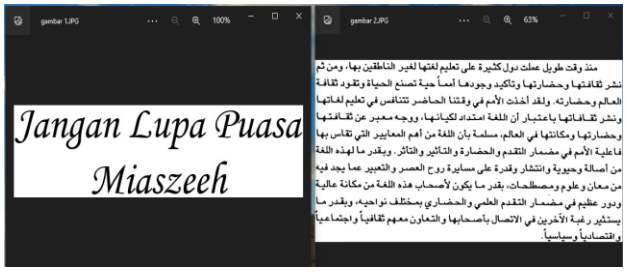
Gambar 5. Rancangan model (3)

Pada penelitian ini secara tidak langsung output yang dihasilkan merupakan hasil pengabungan dari metode OCR, VQ-VAE dan gTTS yang dilakukan sebagai upaya untuk menghasilkan kualitas suara yang lebih baik dan natural serta mempertahankan informasi asli dengan alur proses sebagaimana yang ditunjukkan pada gambar 5 diatas. Pengujian yang dilakukan dalam penelitian ini akan berfokus pada upaya untuk mengukur kualitas suara yang dihasilkan seperti kejelasan, keakuratan dan kesesuaian suara dengan teks yang dikenali, dimana sumber teks berasal dari gambar teks dengan berbagai jenis font, ukuran teks, dan background yang diperggunakan.

F. Dataset

Dalam penelitian ini, dataset yang kami perggunakan berupa gambar teks dengan mencakup berbagai jenis font dan ukuran teks. Dataset yang dimiliki diusahakan juga mencakup penggunaan bahasa yang berbeda, dengan tujuan untuk mengetahui akurasi pengubah suara disamping juga akurasi pengenalan teks.

P-ISSN: 2302-3295



Gambar 6. Contoh dataset yang digunakan

Sebelum diperggunakan dalam deteksi karakter, dataset yang diperggunakan melalui proses *preprocessing* dengan menghilangkan *noise*/noda pada gambar, meningkatkan ketajaman gambar dan memperjelas teks. Proses ini perlu dilakukan untuk meningkatkan kualitas dan kejelasan teks yang diekstraksi.

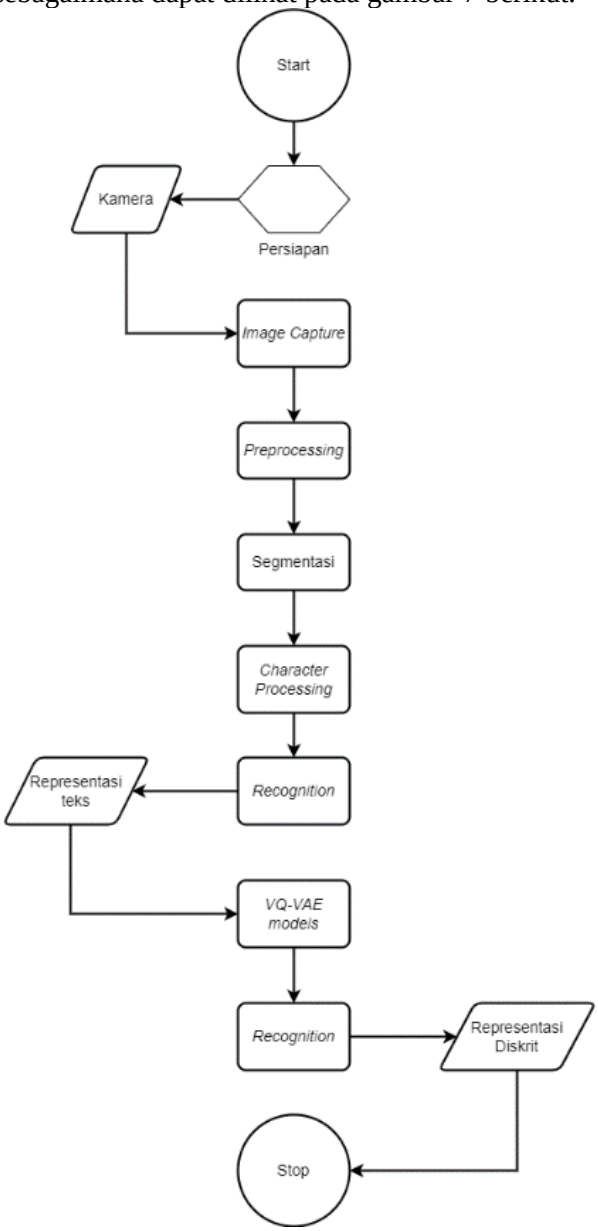
G. Ekstraksi Teks

Dataset yang telah melalui proses preprocessing selanjutnya akan diekstraksi teks yang terdapat didalamnya. Mekanisme ekstraksi dilakukan dengan memanfaatkan *Optical Character Recognition* (OCR), dimana OCR akan dilatih menggunakan dataset gambar teks yang telah disesuaikan.

Beberapa tahapan dilakukan dalam proses ekstraksi teks ini, sebagai berikut:

1. Persiapan dataset gambar, dalam tahapan ini dipastikan dataset gambar yang diperggunakan sudah melalui proses preprocessing sehingga dataset yang diperggunakan memiliki kualitas gambar dan teks yang terlihat jelas.
2. Segmentasi teks, tahapan ini akan membagi gambar kedalam unit-unit teks terpisah menggunakan mekanisme deteksi tepi dan pemisahan kolom, sehingga dapat dikenali secara individual.
3. Pemrosesan karakter, setelah dilakukan segmentasi tahapan lanjutan adalah melakukan deteksi dan identifikasi terhadap karakter dalam gambar serta mengubah menjadi representasi teks. OCR akan memproses dan mengenali karakter satu per satu sesuai jenis huruf atau karakter didalam gambar. Hasil representasi teks ini akan menjadi inputan yang dianalisa menggunakan metode *Vector Quantized Variational Autoencoder* (VQ-VAE) untuk memperoleh representasi disktrik berupa potongan-potongan teks. Representasi diskrit sebagai hasil proses *encoder-decoder* dari ekstraksi teks diharapkan mampu mempertahankan informasi penting dari teks.
4. Output ekstraksi berupa representasi diskrit akan diperggunakan sebagai data input dalam tahapan konversi ke suara.

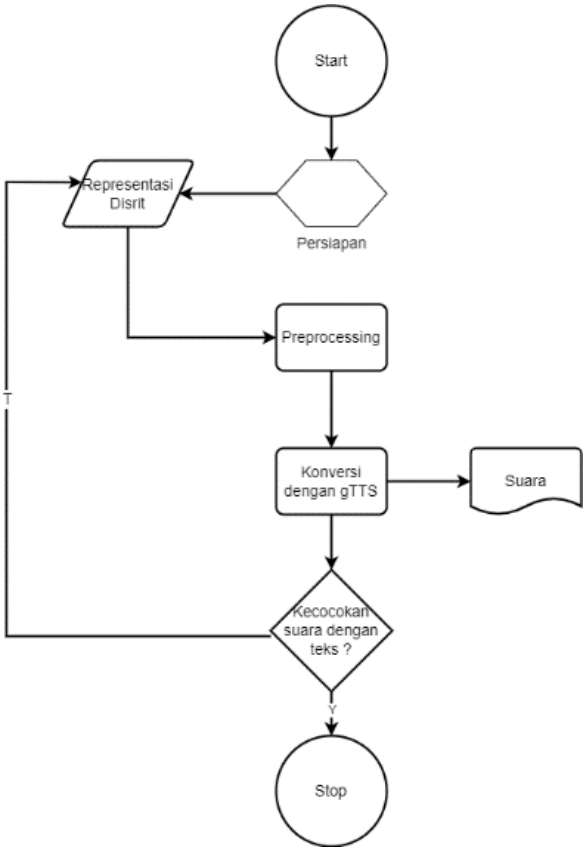
Mekanisme ekstraksi yang dilakukan sebagaimana dapat dilihat pada gambar 7 berikut.



Gambar 7. Alur proses ekstraksi teks

H. Konversi ke Suara

Representasi diskrit sebagai output dari proses ekstraksi teks diperggunakan sebagai data input yang akan dikonversi menjadi suara, dengan menggunakan teknik *Text-to-Speech* dari *google* (gTTS) proses konversi dilakukan. Penggunaan representasi diskrit sebagai data input ditujukan untuk meningkatkan kualitas suara yang dihasilkan dan lebih terkesan alami, disamping juga untuk mempertahankan informasi asli teks. Mekanisme konversi ke suara yang dilakukan dapat dilihat pada gambar 8 berikut.



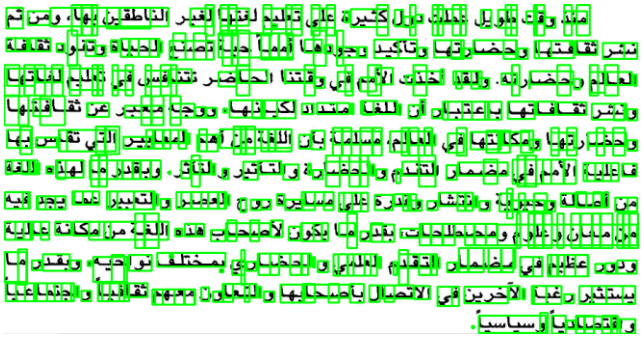
Gambar 8. Alur proses konversi ke suara

I. Pengujian

Pengujian dilakukan untuk mengetahui sejauh mana akurasi antara representasi teks dan suara hasil konversi dengan gambar teks asli, hal ini dilakukan untuk memastikan informasi yang terdapat didalam gambar teks asli tetap ada dan sesuai. Pengujian dilakukan dengan melihat kesesuaian teks representasi diskri dan kecocokan suara hasil konversi dengan teks pada gambar teks asli. Indikator yang diperggunakan seperti jumlah karakter yang berhasil diekstraksi, kesesuaian teks atau durasi waktu suara hasil konversi.

III. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dalam penelitian ini diawali dengan proses pendeteksian karakter dari gambar data input yang telah melalui proses *preprocessing*, sebagaimana yang dapat dilihat pada gambar 9 berikut.

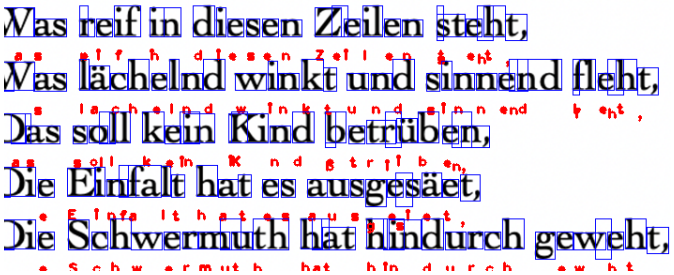


Gambar 9. Deteksi karakter

Seperti yang ditunjukkan pada gambar 9 diatas, masing-masing karakter dari teks yang terdapat pada gambar akan dideteksi dan ditandai menggunakan deteksi tepi. Proses ekstraksi dilakukan dengan memanfaatkan *Optical Character Recognition* (OCR) sebagai model yang telah terlatih untuk melakukan deteksi dan ekstraksi teks dari gambar. Hasil deteksi yang dilakukan selanjutnya akan diekstraksi menjadi teks yang dapat disimpan dan diedit seperti berikut:

منذ وقت طويل عملت دول كثيرة على
تعليم لغتها لغير الناطقين بها ومن
ثمنشر ثقافتها وحضارتها وتأكيد
وجودها أممأ حية تصنع الحياة وتقود
ثقافة العالم وحضارته .ولقد أخذت الأمم
في وقتنا الحاضر تتنافس في تعليم
لقاتهاونشر ثقافاتها باعتبار أن
اللغة امتداد لكيانهاء ووجه معبر عن
.ثقافتهاوحضارتها ومكانتها في العالم
مسلمة بأن اللغة من أهم المعايير
التي تقاس بهاقاعلية الأمم في مضمار
.التقدم والحضارة و التأثير والتأثر
وبقدر ما لهذه اللغةمن أصالة وحيوية
وانتشار وقدرة على مسايرة روح العصر
والتعبير عما يجد فيهمن معان وعلوم
ومصطلحات ,بقدر ما يكون لأصحاب هذه
اللغة من مكانة عالية
ودور عظيم في مضمار التقدم العلمي و
الحضاري بمختلف نواحيه، وبقدر نا
يستثير رغبة الآخرين في الاتصال
بأصحابها والتعاون معهم ثقافيا
.واجتماعياواقتصاديا وسياسيا

Selanjutnya teks hasil ekstraksi diproses untuk mendapatkan representasi diskrit menggunakan model VQ-VAE, dimana proses ini akan mengekstraksi dan mengenali memproses dan mengenali karakter satu per satu sesuai jenis huruf atau karakter sehingga akan menjadi potongan-potongan teks, sebagaimana yang ditunjukkan gambar 10 berikut.



Gambar 10. Representasi Diskrit

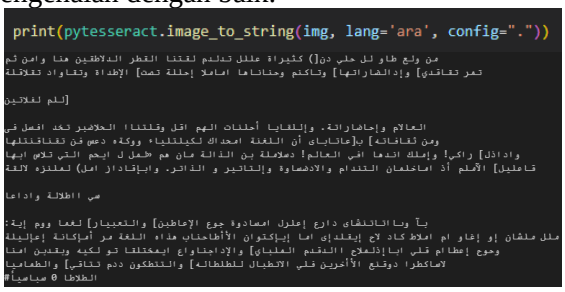
Didalam proses pengubahan menjadi representasi diskrit, tidak semua teks yang terdapat didalam gambar dapat diubah dan disesuaikan menjadi representasi diskrit, terkadang model VQ-VAE akan mengubah huruf atau karakter menjadi karaktek alfabet yang umum dikenali, terlepas teks yang digunakan tidak menggunakan huruf alfabet, hal ini sebagaimana ditunjukkan pada gambar 11 berikut.

P-ISSN: 2302-3295

Gambar 11. Representasi Diskrit yang tidak sesuai

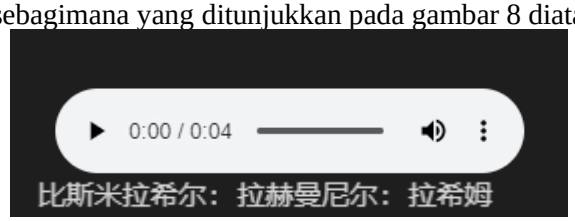
Ketidaksesuaian representasi diskrik yang diperoleh juga dipengaruhi oleh beberapa hal disamping perbedaan jenis karakter, seperti penggunaan rata yang berbeda, dan letak pemisah baris yang berbeda. Imbas dari ketidaksesuaian dalam representasi diskrik yang didapatkan adalah hasil audi yang diperoleh tidak sesuai dengan teks yang diekstraksi.

Untuk mengatasi hal diatas, dilakukan penambahan fitur pengenal bahasa dengan konfigurasi default pada mekanisme konversi gambar ke teks, sehingga mampu melakukan pengenalan dengan baik.



Gambar 12. Fitur pengenalan bahasa dan hasil konversi gambar

Representasi diskrit yang sudah sesuai menjadi data input yang diubah menjadi suara menggunakan *Text-to-Speech* dari *google* (gTTS) dengan proses sebagimana yang ditunjukkan pada gambar 8 diatas.



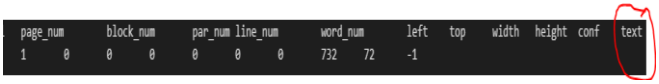
Gambar 13. Contoh hasil ubahan ke suara

Dalam pengujian yang telah dilakukan, terdapat beberapa temuan, dengan berfokus pada akurasi antara representasi teks dan suara hasil konversi dengan gambar teks asli dan berdasar pada informasi asli berupa teks “bismillahirrahmanirrahim”, telah diperoleh hasil pengujian seperti berikut.

Tabel 1. Pengujian Akurasi

Gambar teks	Teks ekstraksi	Bahasa	Durasi audio
بسم الله الرحمن الرحيم	بسم الله الرحمن الرحيم	Arabic	0:02
ቢስሚላሂር ራሀማኒር ረሂም	ቢስሚላሂር ራሀማኒር ረሂም	Amharic	0:02
బిస్మిల్లాహిర్ రహ్మానిర్ రహీమ్	బిస్మిల్లాహిర్ రహ్మానిర్ రహీమ్	Telugu	0:02
比斯米拉希尔: 拉赫曼尼尔: 拉希姆	比斯米拉希尔: 拉赫曼尼尔: 拉希姆	Chinesse	0:04
비스밀라히르 라흐마니르 라힘	-	Korea	0:02
ബിസ്മില്ലാഹിർ റഹ്മാനിർ റഹീം	ബിസ്മില്ലാഹിർ റഹ്മാനിർ റഹീം	Malayalam	0:02
ビスミラヒル・ラフマニル・ラヒム	ビスミラヒル・ラフマニル・ラヒム	Japan	0:02
бісміллахір рахманір Рахім	бісміллахір рахманір Рахім	Ukraina	0:02
बिस्मिल्लाहिर रहमनीर रहीम	बिस्मिल्लाहिर रहमनीर रहीम	Sankrit	-
μπισμιλαχίρ ραχμανίρ Ραχίμ	μπισμιλαχίρ ραχμανίρ Ραχίμ	Grecce	0:02
बिस्मिल्लाहिर रहमानिर रहीम	बिस्मिल्लाहिर रहमानिर रहीम	Hindi	0:02
บิสมิลลาฮิรเราะห์มานิรรอฮีม	บิสมิลลาฮิรเราะห์มานิรรอฮีม	Thai	0:03

Berdasar hasil pengujian pada tabel 1 diatas, diketahui bahwa ada beberapa jenis huruf yang belum dapat dilakukan ekstraksi teks maupun konversi suara. Pada kasus teks pada bahasa korea, ekstraksi teks tidak dapat diperoleh dikarenakan kolom teks dalam *shape image* yang tidak terdeteksi, sedangkan dalam bahasa sankrit belum mampu dilakukan pengubahan disebabkan data bahasa sankrit belum didigitalkan sehingga hanya dapat diekstraksi teksnya. Dalam pengujian yang dilaksanakan diperoleh sebanyak 10 atau 83,33% data uji yang berhasil untuk diekstraksi teks, dilakukan konversi representasi dan diubah ke data suara.



Gambar 14. Kolom teks pada bahasa korea kosong

IV. KESIMPULAN

Dalam penelitian yang telah dilaksanakan, penelitian memperoleh beberapa kesimpulan sebagai berikut:

a) Pendekatan yang diperggunakan dalam penelitian ini yaitu menggunakan *Optical Character Recognition* dengan konversi representasi *Vector Quantized Variational Autoencoder* dengan pengubah suara *Text-to-Speech* dari *google* (gTTS) cukup efektif dalam mempertahankan informasi asli dan menghasilkan suara natural.
b) Hasil pengujian diperoleh akurasi konversi dan pengubahan sebanyak 83,33% dengan 10 data uji dapat dikonversi dan diubah dengan baik. Kedepannya perlu adanya *preprocessing* dataset yang lebih ketat sehingga dapat diperoleh hasil uji yang lebih baik.

DAFTAR PUSTAKA

1. Hsu WN, Harwath D, Miller T, Song C, Glass J. Text-Free Image-to-Speech Synthesis Using Learned Segmental Units [Internet]. Available from: https://wnhsu.github.io/image-to-speech-demo.
2. Anuradha I, Liyanage C, Wijayawardhana H, Weerasinghe R. Deep learning based sinhala Optical Character Recognition (OCR). In: 20th International Conference on Advances in ICT for Emerging Regions, ICTer 2020 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 298–9.
3. Effendi J, Sakti S, Nakamura S. End-to-End Image-to-Speech Generation for Untranscribed Unknown Languages. IEEE Access. 2021;9:55144–54.
4. Sahlol AT, Abd Elaziz M, Al-Qaness MAA, Kim S. Handwritten arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set. IEEE Access. 2020;8:23011–21.
5. Institute of Electrical and Electronics Engineers. SSD'19 : the 16th International Multiconference on Systems, Signals & Devices : March 21-24, 2019, Istanbul, Turkey.
6. Abdu S, Saranga' JL, Sulu V, Wahyuni R. DAMPAK PENGGUNAAN GADGET TERHADAP PENURUNAN KETAJAMAN PENGLIHATAN. Jurnal Keperawatan Florence Nightingale. 2021 Jun 26;4(1):24–30.
7. Dwi Hasriani R, Pencegahan dan Pengendalian Penyakit Tidak Menular D, Kesehatan KR. 645 HIGEIA 4 (4) (2020) HIGEIA JOURNAL OF PUBLIC HEALTH RESEARCH AND DEVELOPMENT Hipertensi dengan Katarak pada Peserta Skrining Gangguan Penglihatan. 2020; Available from: http://journal.unnes.ac.id/sju/index.php/higeiahttps://doi.org/10.15294/higeia/v4i4/38745
8. Edukasi pencegahan penyakit mata.
9. Larsson A, Segerås T. Automated invoice handling with machine learning and OCR Automatiserad fakturahantering med maskininlärning och OCR. DEGREE PROJECT COMPUTER ENGINEERING. 2016.
10. Alghyaline S. Arabic Optical Character Recognition: A Review. Vol. 135, CMES - Computer Modeling in Engineering and Sciences. Tech Science Press; 2023. p. 1825–61.
11. Prayogi YR, Budiman SN. Color Grading Systems to Classify Ripeness of Apple Mango Fruit. Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi. 2018 Oct 3;3(2):57–61.
12. Firdaus A, Syamsu Kurnia M, Shafera T, Firdaus WI, Teknik J, Politeknik K, et al. Implementasi Optical Character Recognition (OCR) Pada Masa Pandemi Covid-19 *1. Vol. 13, Jurnal JUPITER. 2021.
13. Niharika GL, Bano S, Kumar PS, Deepika T, Thumati H. Character Recognition using Tesseract enabling Multilingualism. In: Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 1321–7.
14. Girin L, Leglaive S, Bie X, Diard J, Hueber T, Alameda-Pineda X. Dynamical variational autoencoders: A comprehensive review. Vol. 15, Foundations and Trends in Machine Learning. Now Publishers Inc; 2021. p. 1–175.
15. van den Oord DeepMind A, Vinyals DeepMind O, Kavukcuoglu DeepMind K. Neural Discrete Representation Learning.
16. Tjandra A, Sisman B, Zhang M, Sakti S, Li H, Nakamura S. VQVAE Unsupervised Unit Discovery and Multi-scale Code2Spec Inverter for Zerospeech Challenge 2019. 2019 May 27; Available from: http://arxiv.org/abs/1905.11449
17. Khete T, Bakshi A. Autonomous Assistance System for Visually Impaired using Tesseract OCR & gTTS. In: Journal of Physics: Conference Series. Institute of Physics; 2022.
18. Karmel A, Sharma A, Pandya M, Garg D. IoT based Assistive Device for Deaf, Dumb and Blind People. In: Procedia Computer Science. Elsevier B.V.; 2019. p. 259–69.
19. Fahn CS, Chen SC, Wu PY, Chu TL, Li CH, Hsu DQ, et al. Image and Speech Recognition Technology in the Development of an Elderly Care Robot: Practical Issues Review and Improvement Strategies. Healthcare (Switzerland). 2022 Nov 1;10(11).
20. https://jdih.kemenkeu.go.id/fulltext/2008/14TAHUN2008UUPenjel.htm#:~:text=Dalam%20Undang%2DUndang%20Dasar%20Negara,Informasi%20dengan%20menggunakan%20segala%20jenis diakses pada 4 Juli 2023
21. https://www.alomedika.com/prevalensi-dan-penyebab-gangguan-tajam-penglihatan-pada-populasi-di-asia-tenggara diakses pada 6 Juli 2023
22. https://nasional.kompas.com/read/2022/10/04/19365681/perdami-80-persen-gangguan-penglihatan-di-indonesia-mestinya-bisa-ditangani#:~:text=Adapun%20sejauh%20ini%2C%20terdapat%208,gangguan%20penglihatan%20sedang%20dan%20berat. diakses pada 10 Juli 2023

P-ISSN: 2302-3295

23. https://man2kotapayakumbuh.sch.id/2020/03/03/tuna-netra-keutamaan-dan-balasan-surga-untuknya/ diakses pada 12 Juli 2023
24. Kementrian Kesehatan RI. (2018). InfoDATIN Pusat Data Informasi Kesehatan RI. Jakarta: KEMENKES RI. diakses pada 12 Juli 2023
25. Kemal, David dkk. 2018. Optical Character Recognition (OCR) menggunakan Tesseract dan Penerapannya pada Industri Digital di Indonesia. https://mti.binus.ac.id/2018/12/26/optical-character-recognition-ocr-menggunakan- diakses pada 13 Juli 2023
26. tesseract-dan-penerapannya-pada-industri-digital-di-indonesia/, diakses pada 13 Juli 2023
27. https://pypi.org/project/gTTS/ diakses pada 14 Juli 2023