

Exploring Image-to-Text Methods for Molecular Representation and Relative Molecular Mass Computation

Ahmad Khozi Waridi

Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
230605110083@student.uin-malang.ac.id

October 14, 2024

Abstract

Berbagai aplikasi teknologi, termasuk pengenalan karakter optik (OCR), membutuhkan pemrosesan gambar. Studi ini mengkaji teknik image-to-text yang menggunakan OCR untuk molekul relative dan perhitungan massa molekul relative. Fokus penelitian adalah mengubah gambar atau senyawa kimia menjadi format teks untuk perhitungan molekul relative. Studi ini mengevaluasi berbagai algoritma OCR untuk mengenali karakter dan simbol kimia dari gambar senyawa. Metode yang efektif dikembangkan untuk mengonversi hasil OCR ke dalam format teks yang sesuai dengan standar kimia, dan algoritma digunakan untuk menghitung massa molekul relative dari teks yang dihasilkan. Hasil penelitian menunjukkan bahwa teknik OCR yang disarankan dapat mengenali dan menerjemahkan karakter kimia dengan cukup akurat, namun saat pengenalan senyawa kimia program tidak mampu mengenali secara akurat. Sehingga memberikan nilai molekul relative yang kurang akurat dan bahkan salah.

Kata kunci: OCR, molekul relative, massa molekul relative, image-to-text, pemrosesan citra.

1 Pendahuluan

Dalam era digital yang semakin berkembang, pemrosesan citra telah menjadi salah satu bidang yang sering di aplikasikan pada teknologi. Salah satu aplikasi utama dari pemrosesan citra adalah Optical Character Recognition (OCR), sebuah teknologi yang mampu mengonversi gambar teks menjadi teks yang dapat diolah kemudian.

Teknologi ini memiliki potensi besar dalam berbagai bidang, termasuk pengenalan karakter digitalisasi dokumen, dan analisis gambar ilmiah. Pada penelitian kali ini saya berfokus pada metode image-to-text menggunakan OCR untuk pengenalan gambar senyawa dan perhitungan molekul relative. Penggunaan molekul relative sangat penting dalam dalam berbagai bidang seperti kimia, biologi, dan farmasi. Sebab molekul realtive merupakan salah satu dasar untuk perhitungan senyaw-senyawa selanjutnya.

Salah satu tantangan utama dalam konversi gambar molekul ke teks adalah akurasi dan efisiensi dari teknologi OCR dalam mengenali dan menerjemahkan karakter dan simbol kimia yang kompleks. Oleh karena itu, penelitian ini mengeksplorasi berbagai metode OCR untuk mengidentifikasi huruf, angka, dan simbol dalam gambar senyawa, serta bagaimana representasi ini dapat digunakan untuk menghitung massa molekul relative dengan tepat. Penelitian ini bertujuan untuk:

1. Mencoba kinerja berbagai algoritma OCR dalam mengenali karakter dan simbol kimia dari gambar senyawa.
2. Mengimplementasikan algoritma untuk menghitung massa molekul relative berdasarkan teks yang dihasilkan dari proses OCR.

Dengan melakukan penelitian ini, saya berharap dapat memberikan kontribusi signifikan dalam bidang pemrosesan citra dan bidang kimia, serta membuka jalan untuk aplikasi lebih lanjut dari teknologi OCR dalam konteks ilmiah dan industri. Hasil dari penelitian ini diharapkan dapat membentuk ide baru dalam pengembangan teknologi terutama teknologi OCR.