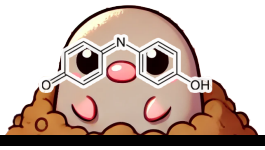


MolMole: Molecule Mining from Scientific Literature

LG AI Research

Abstract

Extracting molecular structures and reaction data from scientific literature remains a significant challenge due to the unstructured nature of chemical information embedded in figures and diagrams. While deep learning has enabled automation, existing approaches often rely on preprocessing steps or focus on isolated elements rather than full-page content, limiting their scalability and efficiency. We introduce MolMole, a vision-based deep learning framework that directly processes page-level molecular data, eliminating the need for external document parsing. Unlike existing systems, MolMole integrates molecule detection, reaction diagram parsing, OCSR, and an interactive editing tool into a single, end-to-end workflow, ensuring efficient and accurate chemical data extraction. Additionally, MolMole achieves higher processing speed and extraction accuracy than publicly available benchmarks, making it well-suited for large-scale data applications. To support robust evaluation, we establish a page-level test set and an evaluation metric designed to assess molecular information extraction at the document level. Through extensive evaluation, MolMole demonstrates superior performance over existing approaches, providing a scalable and efficient solution for molecular data extraction, cheminformatics, and AI-driven chemical research. An instance of the MolMole web application is available at *https://MOLMOLE.ai*.

1 Introduction

The rapid growth of scientific publications in chemistry and materials science has led to an overwhelming accumulation of molecular structure and reaction data. However, much of this valuable information remains embedded in unstructured formats, such as images, figures, and complex diagrams, making automated extraction challenging. Converting this data into machine-readable formats is essential for integrating it into public databases, enabling large-scale analysis, and accelerating research. Traditionally, this extraction process has been manual and time-consuming, requiring significant human effort and financial resources.

In recent years, several AI-driven frameworks have been developed for document-level molecular data extraction, with DECIMER and OpenChemIE being among the most prominent. DECIMER was the first publicly available framework to incorporate molecule segmentation, classification, and Optical Chemical Structure Recognition (OCSR). However, it lacks the ability to process reaction diagrams, limiting comprehensive chemical data extraction. In contrast, OpenChemIE achieves strong OCSR and reaction diagram parsing performance by leveraging multiple AI models. However, it relies on an external layout parser model to crop document elements, which can lead to detection failures in complex layouts.

In this work, we introduce **MolMole**, a vision-based deep learning toolkit designed for page-level molecular information extraction that overcomes the limitations of existing approaches. Unlike previous systems, MolMole integrates molecule detection, reaction diagram parsing, OCSR, and

an interactive molecule editing tool into a unified workflow, processing page-level input directly without relying on external layout parsing. This enables a fully vision-based and seamless solution for extracting chemical information from scientific literature.

The limited advancement in developing a comprehensive framework for molecular information extraction is largely due to the lack of benchmark datasets for evaluating the entire extraction process. While numerous models exist for individual tasks such as OCSR, molecule segmentation, and reaction diagram parsing, their true value emerges only when integrated into a complete system. Therefore, establishing such a benchmark is essential.

Beyond its comprehensive functionality, MolMole is designed for efficiency and scalability, achieving faster processing speeds than publicly available benchmark systems while delivering superior performance, making it highly suitable for handling large-scale datasets. To ensure rigorous evaluation, we curated a page-level test set and introduced an evaluation metric tailored for assessing molecular information extraction at the document level.

In our experiments, MolMole consistently outperformed existing toolkits, including OpenChemIE, DECIMER, and ReactionDataExtractor 2.0, achieving high F1 scores in both molecular structure extraction and reaction diagram parsing. Additionally, on an open benchmark dataset for OCSR, MolMole demonstrated broader molecular coverage and achieved accuracy on par with or exceeding state-of-the-art solutions. With these advancements, MolMole establishes a robust and efficient framework for automating chemical data extraction, paving the way for large-scale chemical data curation and enabling further research and AI model development in cheminformatics.

The following list summarizes the key contributions of **MolMole**:

- **End-to-End Page-Level Extraction:** Unlike existing systems, MolMole integrates molecule detection, reaction diagram parsing, OCSR, and an interactive molecule editing tool into a single workflow while processing page-level input directly.
- **Page-Level Evaluation Benchmark:** We constructed a page-level test set and introduced an evaluation metric specifically designed to assess molecular information extraction at the document level.
- **High-Speed and High-Accuracy Processing:** MolMole outperforms publicly available benchmark systems in both processing speed and extraction accuracy, making it well-suited for handling large volumes of data efficiently.

2 MolMole Pipeline

2.1 Overview

MolMole offers an intuitive and automated solution for extracting molecular image data from scientific literature while incorporating human oversight to ensure high-quality results. The pipeline begins with a PDF document, which is converted into individual page images for analysis. These images are then processed sequentially by specialized AI models.

The first model, ViDetect, detects molecular structures by identifying and localizing their bounding boxes within the document. Once the molecular regions are identified, ViReact extracts key information from reaction diagrams, including reactants, conditions, and products. Finally, ViMore processes the localized molecular regions, extracting structural details and converting them into machine-readable formats such as molfile or SMILES strings.

Following automated extraction, users can review and refine the results using an integrated editing tool, ensuring any inaccuracies are corrected manually. This hybrid approach combines the efficiency of automation with the precision of human validation.

Once finalized, the extracted data can be saved or exported in formats such as JSON or Excel, facilitating seamless integration into downstream workflows. The following sections provide a detailed breakdown of the AI models and editing tools that power MolMole's capabilities.

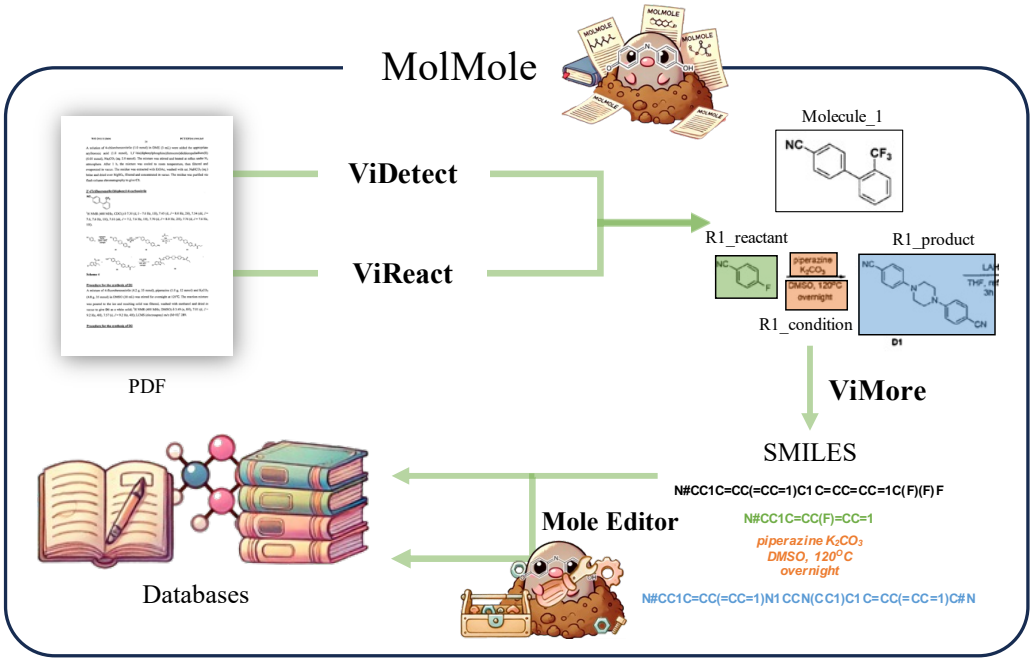


Figure 1: MolMole pipeline. The input PDF is first converted into images. Molecule detection is then performed on the images while reaction parsing runs in parallel. The results from both processes are merged, and non-maximum suppression (NMS) is applied to resolve duplicate detections. Finally, molecule conversion is conducted to generate the final Mol file or SMILES for all detected molecules.

2.2 Models

2.2.1 ViDetect for Molecule Detection

ViDetect (Vision Detection) is an object detection model designed to predict bounding boxes for molecular structures in document images. Its architecture is derived from DINO and is trained end-to-end on our private dataset. To enhance detection accuracy, all predicted bounding boxes undergo post-processing to remove overlapping proposals based on confidence scores and size constraints.

Existing molecule detection models follow different approaches. DECIMER-segmentation, for example, is based on Mask R-CNN, a segmentation model that isolates molecular structures from background noise by segmenting within the bounding box. While this approach helps remove extraneous elements, its high computational cost makes it impractical for large-scale data processing. Additionally, models like MolScribe can effectively handle noisy inputs, making segmentation-based preprocessing unnecessary, as it does not significantly improve accuracy.

Another approach is MolDet, the molecule detection model integrated into OpenChemIE, which follows a sequence-to-sequence (seq2seq) framework. This auto-regressive approach introduces a drawback—inference time increases with the number of molecules in an image, making it less efficient for large-scale applications.

We chose a DETR-based architecture for ViDetect because of its efficiency and strong performance in object detection tasks. DETR offers a balance of speed and accuracy, making it well-suited for processing large volumes of molecular data. As shown in our evaluation section, ViDetect outperforms existing models, achieving the best results across key metrics.

2.2.2 ViReact for Reaction Diagram Parsing

ViReact (Vision Reaction Parsing) is our reaction diagram parsing model, designed to extract structured information directly from page-level document images. It identifies reaction roles, including reactants, conditions, and products, while also predicting bounding box coordinates and entity types

for each component. Unlike models that rely on layout parsers to crop regions of interest before processing, ViReact operates directly on full-page inputs. This is particularly important because layout parsers often fail to detect reaction diagrams accurately, leading to missing or incomplete extractions. For example, in two-column documents, a reaction diagram that spans both columns may be incorrectly segmented, resulting in incomplete data. By eliminating this preprocessing step, ViReact ensures robust and reliable detection of reaction diagrams across diverse document layouts.

To enable effective page-level reaction parsing, we constructed a custom dataset with page-level annotations, allowing the model to detect reaction diagrams without requiring pre-segmentation. This dataset includes reaction diagrams from both articles and patents, covering a wide range of formatting styles and reaction structures. The diversity of this dataset enables ViReact to generalize effectively, making it significantly more adaptable than traditional approaches that rely on layout parsing. ViReact follows an encoder-decoder architecture, inspired by RxnScribe. The encoder abstracts the input image into hidden representations, while the decoder generates the structured reaction sequence in an autoregressive manner.

At inference time, ViReact applies post-processing to refine predictions, correcting errors such as duplicate reactions and empty entities. By parsing reaction diagrams directly from page-level inputs, ViReact eliminates the need for error-prone multi-step processing, simplifying the pipeline while improving recall in identifying reaction structures from scientific literature. This approach reduces complexity while ensuring higher accuracy and efficiency, making it well-suited for large-scale chemical data extraction.

2.2.3 ViMore for Optical Chemical Structure Recognition

ViMore (Vision Molecule Recognition) is an OCSR (Optical Chemical Structure Recognition) model that converts molecular images into machine-readable formats like Molfiles [1] or SMILES [7]. It detects atom regions, recognizes atom characters, and predicts bond types, assembling this information into a structured molecular format through postprocessing. Trained end-to-end on a private dataset, ViMore ensures high accuracy in molecular structure recognition.

Unlike generative models such as MolScribe [5] and Decimer [6], which translate molecular images directly into SMILES sequences, ViMore adopts a detection-based approach. By explicitly predicting atomic and bond information, it eliminates hallucination errors, enhances transparency, and facilitates layout-aware MOL file generation. Additionally, ViMore requires less training data while remaining highly extensible to molecular structures beyond SMILES constraints, including wavy bonds and polymers.

ViMore also offers advanced capabilities: it can recognize polymer structures with bracket notations, detect wavy bonds frequently found in patents (as shown in Figure 3), and preserve the original spatial layout in MOL files for easier verification (illustrated in Figure 2). Furthermore, it provides a confidence score (low, medium, high) to help users assess the reliability of predictions. A screenshot of ViMore's prediction results with confidence scores is shown in Figure 11 and Figure 12.

2.3 Editing Tool

In addition to the three AI models mentioned above, MolMole provides an editing tool for modifying molecular structures. When reconstructing molecular structures, prediction errors typically occur as minor mistakes, such as missing a single atom or bond, rather than large-scale inaccuracies. With our editing tool, users can easily refine and correct the predicted molecular structure to obtain the desired MOL file.

Unlike conventional public MOL file editing tools such as Marvin and ChemDraw, which modify molecules based on existing MOL files, our editing tool allows vision-based modifications. Users can add new bounding boxes for atoms or create new bonds directly using a vision-based approach, ultimately generating the final MOL file. This is possible because ViMore is a vision/detection-based model, making structural modifications more intuitive and user-friendly. A snapshot of the editing tool is shown in Figure 13.

3 Performance

3.1 Benchmark

A key challenge in developing and evaluating page-level extraction from chemical literature is the lack of end-to-end benchmark dataset. While OCSR benchmarks exist, they focus solely on image-to-molecule conversion without evaluating molecule detection, which is critical for page-level performance. To bridge this gap, we constructed a custom dataset that simulates real-world scenarios where an entire PDF serves as input, requiring the extraction of relevant chemical information.

This dataset includes detailed, manually curated annotations for three core tasks: molecule detection, reaction parsing, and molecule conversion. The dataset comprises a total of 550 pages from scientific articles and patents, selected to capture diverse molecular structures, reaction diagrams, and layout variations. Each page has a full annotation of molecular bounding boxes, reaction diagram components (such as reactants, conditions, and products), and corresponding molecular representations in Molfile format, enabling end-to-end evaluation of the whole pipeline. Table 1 shows the curated testset statistics: number of pages, total number of molecules and total number of reactions. Visualization of some testset is in the Appendix.

Table 1: Testset Statistics

Dataset	# Pages	# Molecules	# Reactions
Patents	300	2,482	728
Articles	250	1,415	294

3.2 Evaluation

We evaluated MolMole mainly against two state-of-the-art chemical information extraction frameworks, DECIMER 2.0 [6] and OpenChemIE [2], both of which offer end-to-end processing from PDFs to extracted data. Specifically, ViDetect is compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect, ViMore with DECIMER's Decimer Transformer and OpenChemIE's MolScribe, and ViReact with OpenChemIE's RxnScribe and ReactionDataExtractor 2.0 [8].

3.2.1 Page-level Molecule Detection and Recognition

Need to add results with explanation.

This section presents the page-level evaluation results, encompassing three distinct assessments: (1) Molecule Detection performance, (2) Molecule Conversion performance using Ground Truth (GT) bounding boxes, and (3) the combined performance of Molecule Detection and Molecule Conversion. The first two evaluations (1) and (2) are conducted independently to assess the effectiveness of molecule detection and conversion separately, without being influenced by each other's outcomes. In contrast, the third evaluation (3) aims to measure the overall performance of the entire pipeline, from molecule detection to conversion.

First, Table 2 summarizes the molecule detection performance of DECIMER Segmentation, MolDetect, and ViDetect. The evaluation considers precision, recall, and F1 score, which are standard object detection metrics, and is conducted on both the Patents and Articles test sets. Each model is evaluated using its default confidence threshold, while () is used for ViDetect.

The result shows ....

Second, Table 3 presents the molecule conversion performance of DECIMER Image Transformer, MolScribe, and ViMore. To evaluate molecule conversion performance in isolation, molecular regions are extracted from the pages of Patents and Articles using ground truth bounding boxes. The predicted molfile is then compared with the ground truth molfile using SMILES matching accuracy and Tanimoto similarity.

The result shows ViMore has the highest score on both patents and journal ...

Table 2: Molecule Detection Performance on Patents and Articles test sets.

	Patents			Articles		
Models	Precision	Recall	F1	Precision	Recall	F1
DECIMER Segmentation [6]	**0.891**	**0.930**	**0.910**	0.839	0.896	0.867
MolDetect [2]	0.796	0.841	0.818	0.764	0.820	0.791
ViDetect (Ours)	0.873	0.900	0.887	**0.874**	**0.903**	**0.888**

Table 3: Molecule Conversion Performance on Patents and Articles test sets.

	Patents		Articles	
Models	SMILES	Tanimoto	SMILES	Tanimoto
DECIMER Image Transformer [6]	.753	.914	.681	.892
MolScribe [5]	.709	.913	.729	.951
ViMore (Ours)	**.909**	**.958**	**.924**	**.967**

Finally, Table 4 presents the overall performance of the full pipeline, from molecule detection to conversion. To assess the combined performance, we modify the conventional object detection metrics by incorporating SMILES string matching into precision and recall.

Given the definitions: $Precision = \frac{TP}{TP+FP}$, $\quad Recall = \frac{TP}{TP+FN}$,

TP, FP, and FN are determined as follows:

$$TP = \sum_{i=1}^{N} \mathbf{1}\left(\max_j IoU(B_{gt}^{(i)}, B_{pred}^{(j)}) \geq \tau \text{ and } SMILES_{gt}^{(i)} = f_{I\to S}(B_{pred}^{(j)})\right) \tag{1}$$

where $B_{gt}^{(i)}$ and $B_{pred}^{(j)}$ denote the i-th ground truth (GT) bounding box and the j-th predicted bounding box, respectively. $SMILES_{gt}^{(i)}$ represents the SMILES string associated with $B_{gt}^{(i)}$, while $f_{I\to S}(B_{pred}^{(j)})$ denotes the predicted SMILES string derived from $B_{pred}^{(j)}$ through the molecular conversion model. The IoU threshold τ is set to 0.5.

A detection is classified as a True Positive (TP) if the predicted bounding box has the highest IoU with a GT bounding box, exceeding the IoU threshold, and the predicted SMILES string matches the GT SMILES string. A False Positive (FP) occurs when a predicted bounding box does not correspond to any GT box or when its associated SMILES string differs from the GT SMILES string, computed as $FP = |B_{pred}| - TP$, where $|B_{pred}|$ represents the total number of predicted bounding boxes. A False Negative (FN) arises when a GT object is not detected by any prediction or when its predicted SMILES string differs from the GT SMILES string, given by $FN = |B_{gt}| - TP$, where $|B_{gt}|$ denotes the total number of ground truth bounding boxes.

The result shows ...

Table 4: Combined performance from molecule detection to conversion on Patents and Articles test sets.

	Patents			Articles		
Models	Precision	Recall	F1	Precision	Recall	F1
DECIMER Segmentation + Image Transformer [6]	.738	.737	.738	.673	.673	.673
MolDetect + MolScribe [2]	.693	.682	.688	.701	.710	.706
ViDetect + ViMore (Ours)	**.888**	**.851**	**.869**	**.911**	**.881**	**.896**

3.2.2 Page-level Reaction Diagram Parsing

To evaluate the performance of our reaction diagram parsing system, we adopt the evaluation methodology proposed in **RxnScribe**. Model performance is measured using *precision*, *recall*, and

F1-score, which quantify how accurately the system identifies reaction components. These metrics are defined as follows:

$$\text{Precision} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}\left(\exists i \in \{1,\ldots,N\} \text{ such that } \tilde{T}_j \text{ matches } T_i\right) \tag{2}$$

$$\text{Recall} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(\exists j \in \{1,\ldots,M\} \text{ such that } T_i \text{ matches } \tilde{T}_j\right) \tag{3}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Precision measures the proportion of correctly predicted reactions out of all predictions made by the model, while recall indicates the proportion of ground truth reactions that were successfully identified. The F1-score provides a balanced assessment by considering both precision and recall. As suggested in **RxnScribe**, we report *micro-averaged precision, recall, and F1 scores* across the entire test set to ensure a comprehensive evaluation.

To apply these metrics, we define the *ground truth reaction set* as:

$$R = \{T_1, T_2, \ldots, T_N\} \tag{5}$$

and the *predicted reaction set* as:

$$S = \{\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_M\} \tag{6}$$

Each predicted reaction $\tilde{T}_j$ is compared with a ground truth reaction T_i by determining their *bounding box overlap*, measured using *Intersection over Union (IoU)*. If the highest IoU score exceeds 0.5, the bounding boxes are considered a successful match.

Evaluation is conducted using two different approaches. The *relaxed matching (soft match)* method considers only molecular entities, ignoring text labels and not distinguishing between reactants and reagents. This approach accounts for cases where molecules positioned near reaction arrows may be visually ambiguous but still play a role in the reaction. The *strict matching (hard match)* method, on the other hand, requires that all reaction components—including reactants, reagents (conditions), and products—are correctly identified. Any misclassification results in an incorrect match.

Table 5 shows the performance of ReactionDataExtractor 2.0, RxnScribe and ViReact.

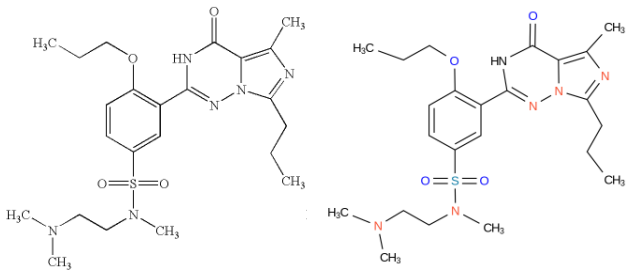
Table 5: Reaction Parsing Performance on our private test set.

Models	Patents						Articles					
	Precision		Recall		F1		Precision		Recall		F1	
	Soft	Hard	Soft	Hard	Soft	Hard	Soft	Hard	Soft	Hard	Soft	Hard
ReactionDataExtractor2.0(w/o LP) [8]	-	-	-	-	-	-	-	-	-	-	-	-
ReactionDataExtractor2.0(w/ LP) [8]	-	-	-	-	-	-	-	-	-	-	-	-
RxnScribe(w/o LP) [4]	0.826	0.496	0.817	0.489	0.822	0.490	0.856	0.525	0.803	0.497	0.829	0.510
RxnScribe(w/ LP) [4]	0.818	0.549	0.691	0.464	0.749	0.503	0.853	0.578	0.721	0.493	0.781	0.532
ViReact (Ours)	**0.968**	**0.898**	**0.923**	**0.856**	**0.945**	**0.876**	**0.940**	**0.826**	**0.956**	**0.840**	**0.948**	**0.833**

3.2.3 OCSR Public Benchmark Evaluation

Need to add results with explanation.

This section evaluates molecule conversion models using publicly available OCSR benchmarks. We compare our model with state-of-the-art methods—DECIMER Image Transformer, MolScribe, and MolGrapher—by running experiments on four standard benchmark datasets: USPTO, Maybridge UoB, CLEF, and JPO, which contain 5719, 5740, 992, and 450 images, respectively.



Input Image Layout-preserving Reconstruction

Figure 2: ViMore prediction preserves the image layout information in the resulting Mol file.

The performance of each method is measured by calculating accuracy as the percentage of molecules correctly recognized. This is determined by comparing the predicted molecules to the ground-truth using both InChI keys and SMILES representations. In addition, the Tanimoto similarity between SMILES strings is computed. The results are presented in Table 6.

The result shows...

Table 6: OCSR performance on public benchmarks. InChI and SMILES refer to the exact matching accuracy for InChI keys and SMILES strings, while Tani represents the Tanimoto similarity.

	CLEF			JPO			UOB			USPTO		
Models	Inchi	SMILES	Tani.	Inchi	SMILES	Tani.	Inchi	SMILES	Tani.	Inchi	SMILES	Tani.
DECIMER I. T. [6]	.720	.715	.848	.664	.667	.846	**.987**	**.901**	**.962**	.630	.608	.921
MolScribe [5]	.796	.830	**.943**	.753	.756	.884	.983	.896	.960	**.934**	**.935**	**.987**
MolGrapher [3]	.496	.493	.905	.556	.560	.783	.950	.869	.935	.639	.635	.944
ViMore (Ours)	**.827**	**.859**	.928	**.804**	**.804**	**.885**	.955	.869	.942	.902	.902	.949

3.3 Inference Speed

BOTH

XXXX Random Paragraph: We compared EXAMORE with two other chemical information extraction frameworks, DECIMER 2.0 and OpenChemIE. These frameworks were selected because they offer end-to-end processing from PDFs to extracted data while achieving state-of-the-art performance. Specifically, ViDetect was compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect. Likewise, ViMore was assessed against DECIMER's Decimer Transformer and OpenChemIE's MolScribe. Finally, ViReact was compared with OpenChemIE's RxnScribe and ReactionDataExtractor2.0. Table 7 shows the inference speed.

Table 7: Inference Speed.

Models	Mol Detection	Reaction	Mol Conversion	Total
DECIMER 2.0 [6]	-	-	0.96 s/it	-
OpenChemIE [2]	-	-	0.62 s/it	-
MolMole (Ours)	-	-	0.41 s/it	-

4 Discussion

MolScribe VS. ViMore

Layout-preserving MOL **NEED JIYEH'S WRITING**

The layout-preserving MOL feature reconstructs a MOL file that retains the original image's structure. Existing OCSR models struggle with this—Decimer lacks atomic position data, and

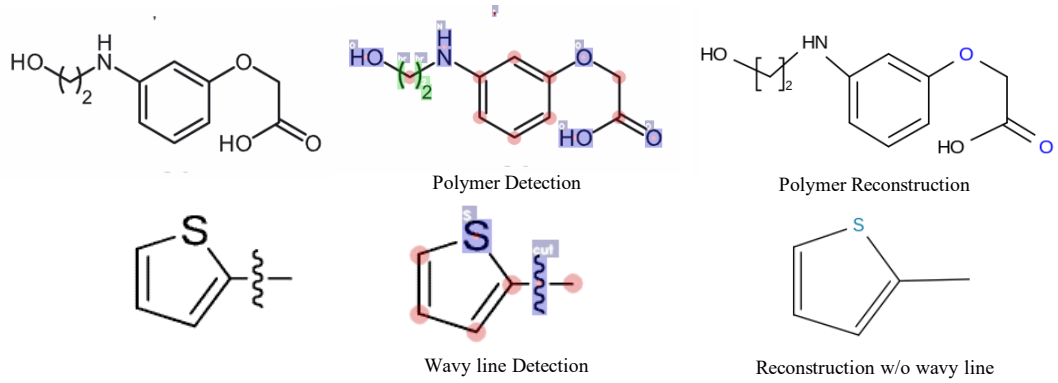


Figure 3: ViMore prediction results of Polymer and Wavy line.

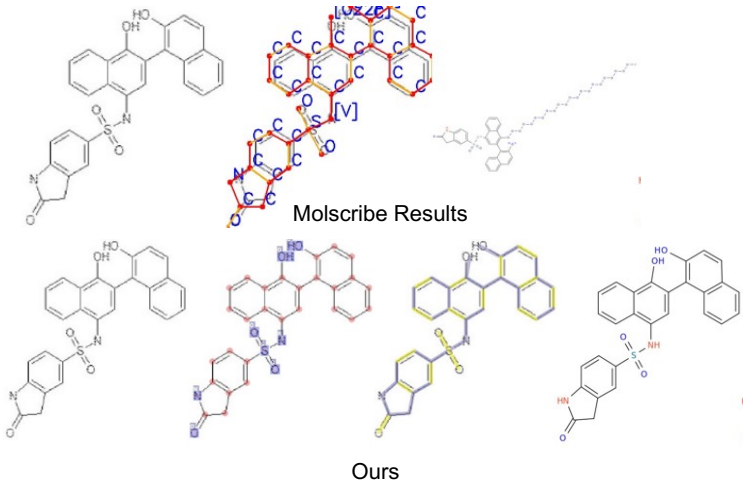


Figure 4: Comparison between MolScribe and ViMore results

MolScribe's coordinates are often inaccurate. In contrast, ViMore, as a detection-based model, generates precise MOL files that match the original layout. As shown in Figure 2, this makes verification easier and allows for quick modifications if needed.

Polymer and Wavy line NEED JIYEH'S WRITING

Polymer structures in molecules are represented by brackets and an adjacent number, indicating that the enclosed structure repeats that many times. Existing OCSR models struggle to accurately recognize this structure, but ViMore can detect both the brackets and numbers and ensure correct conversion. In patents, a wavy line indicates that a different substructure can be attached at that position in the molecular structure. Existing models often misinterpret wavy lines as single bonds or other bond types. In contrast, our model correctly recognizes wavy lines as distinct objects and can generate a MOL file with the wavy line removed, ensuring accurate structure representation. The results are shown in Figure 3.

Reaction Diagram Parsing NEED SEHYUN'S WRITING

XXXX Random Paragraph: We compared EXAMORE with two other chemical information extraction frameworks, DECIMER 2.0 and OpenChemIE. These frameworks were selected because they offer end-to-end processing from PDFs to extracted data while achieving state-of-the-art performance. Specifically, ViDetect was compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect. Likewise, ViMore was assessed against DECIMER's Decimer Transformer

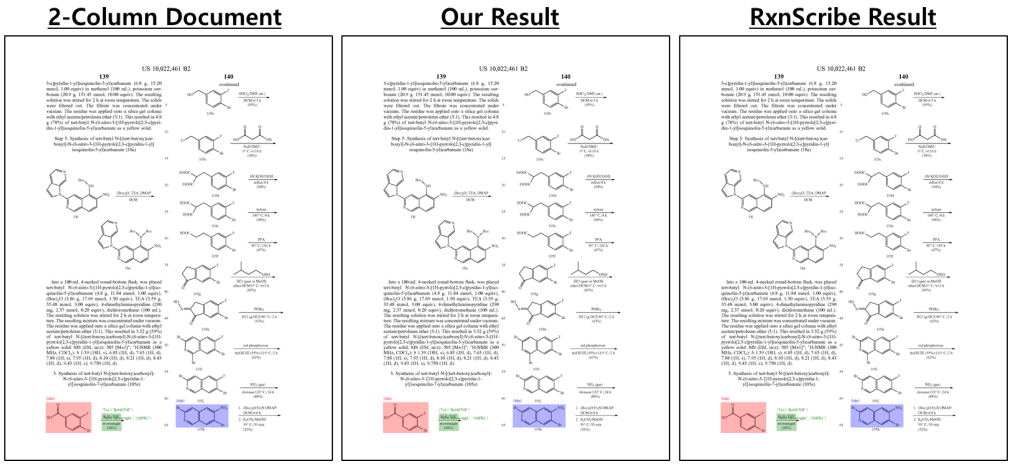


Figure 5: Error cases from Reaction Parsing

and OpenChemIE's MolScribe. Finally, ViReact was compared with OpenChemIE's RxnScribe and ReactionDataExtractor2.0.

5 Conclusion

MolMole offers an effective and scalable solution for automating the extraction of molecular information from scientific literature. Unlike existing tools that require document layout parsing or focus solely on isolated figures, MolMole processes entire page-level document images directly. It leverages specialized AI models for molecule detection, reaction diagram parsing, and molecular structure conversion, streamlining chemical data extraction.

Our evaluation demonstrates that MolMole consistently outperforms existing approaches, achieving high accuracy in extracting molecular structures and reaction diagrams. Additionally, its integrated editing tool enables researchers to refine and validate extracted data. This ensures its suitability for chemical database construction, AI model training, and large-scale cheminformatics applications.

Beyond its technical contributions, MolMole advances cheminformatics by enabling high-throughput, scalable molecular data extraction. Its seamless integration into existing workflows makes it a valuable tool for researchers working with machine-readable chemical data.

Looking ahead, we aim to enhance MolMole's capabilities by improving its robustness in handling complex molecular representations and expanding dataset coverage. Future developments will also focus on integrating MolMole into larger chemical informatics pipelines. As the need for automated molecular data extraction continues to grow, MolMole is well-positioned to play a key role in accelerating AI-driven discoveries in chemistry.

6 Appendix

6.0.1 Contributors

All authors are listed in alphabetical order by last name.

Core Contributors: Eunbi Choi, Seokhee Hong, Junwon Hwang, Hyojin Jeon, Hyunjik Jo, Joonkee Kim, Seonghwan Kim, Soyeon Kim, Sunkyoung Kim, Yireun Kim, Haeju Lee, Jinsik Lee, Kyungmin Lee, Sangha Park, Heuiyeen Yeen, Hyeongu Yun

Contributors: Soyoung An, Kyunghoon Bae, Stanley Jungkyu Choi, Yemuk Choi, Yeonjung Hong, Gerrard Jeongwon Jo, Jiyeon Jung, Yountae Jung, Euisoon Kim, Hyosang Kim, Youchul Kim, Edward Hwayoung Lee, Honglak Lee, Moontae Lee, Seungjun Lee, Woohyung Lim, Sooyoun Park, Yongmin Park, Sihoon Yang, Kyungjae Yoo

6.0.2 Qualitative Results - Hard Case

XXXX Random Paragraph: We compared EXAMORE with two other chemical information extraction frameworks, DECIMER 2.0 and OpenChemIE. These frameworks were selected because they offer end-to-end processing from PDFs to extracted data while achieving state-of-the-art performance. Specifically, ViDetect was compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect. Likewise, ViMore was assessed against DECIMER's Decimer Transformer and OpenChemIE's MolScribe. Finally, ViReact was compared with OpenChemIE's RxnScribe and ReactionDataExtractor2.0.

6.0.3 MolMole Workflow

XXXX Random Paragraph: We compared EXAMORE with two other chemical information extraction frameworks, DECIMER 2.0 and OpenChemIE. These frameworks were selected because they offer end-to-end processing from PDFs to extracted data while achieving state-of-the-art performance. Specifically, ViDetect was compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect. Likewise, ViMore was assessed against DECIMER's Decimer Transformer and OpenChemIE's MolScribe. Finally, ViReact was compared with OpenChemIE's RxnScribe and ReactionDataExtractor2.0.

XXXX Random Paragraph: We compared EXAMORE with two other chemical information extraction frameworks, DECIMER 2.0 and OpenChemIE. These frameworks were selected because they offer end-to-end processing from PDFs to extracted data while achieving state-of-the-art performance. Specifically, ViDetect was compared with DECIMER's Decimer Segmentation and OpenChemIE's MolDetect. Likewise, ViMore was assessed against DECIMER's Decimer Transformer and OpenChemIE's MolScribe. Finally, ViReact was compared with OpenChemIE's RxnScribe and ReactionDataExtractor2.0.

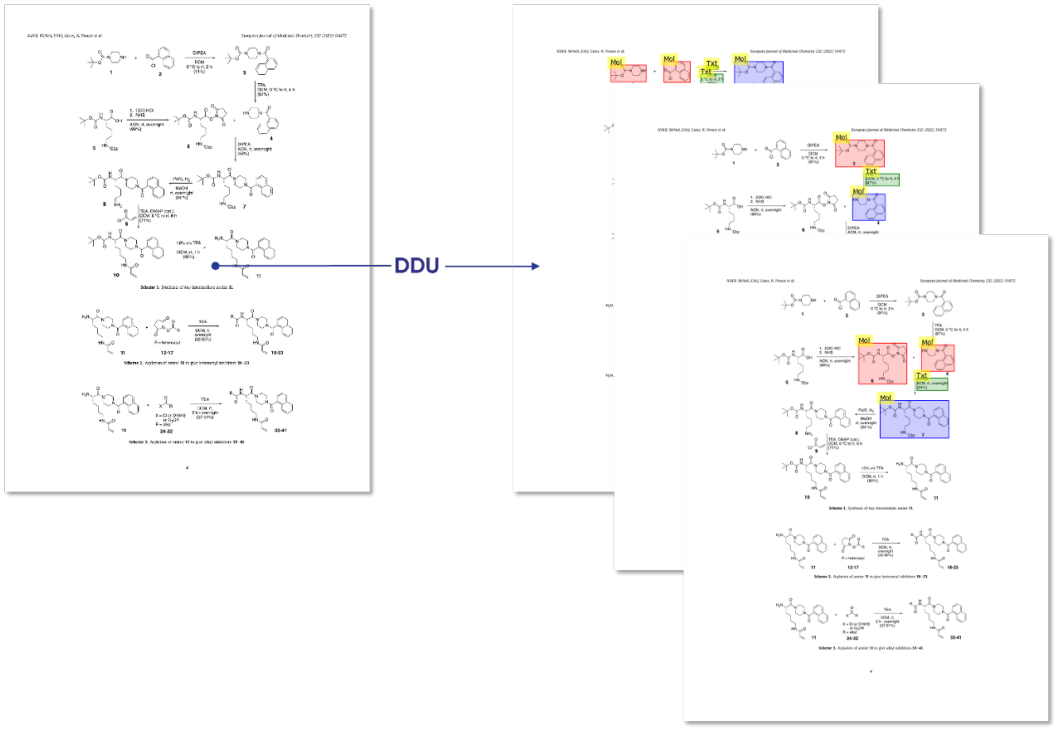


Figure 6: Expected Results/visualization

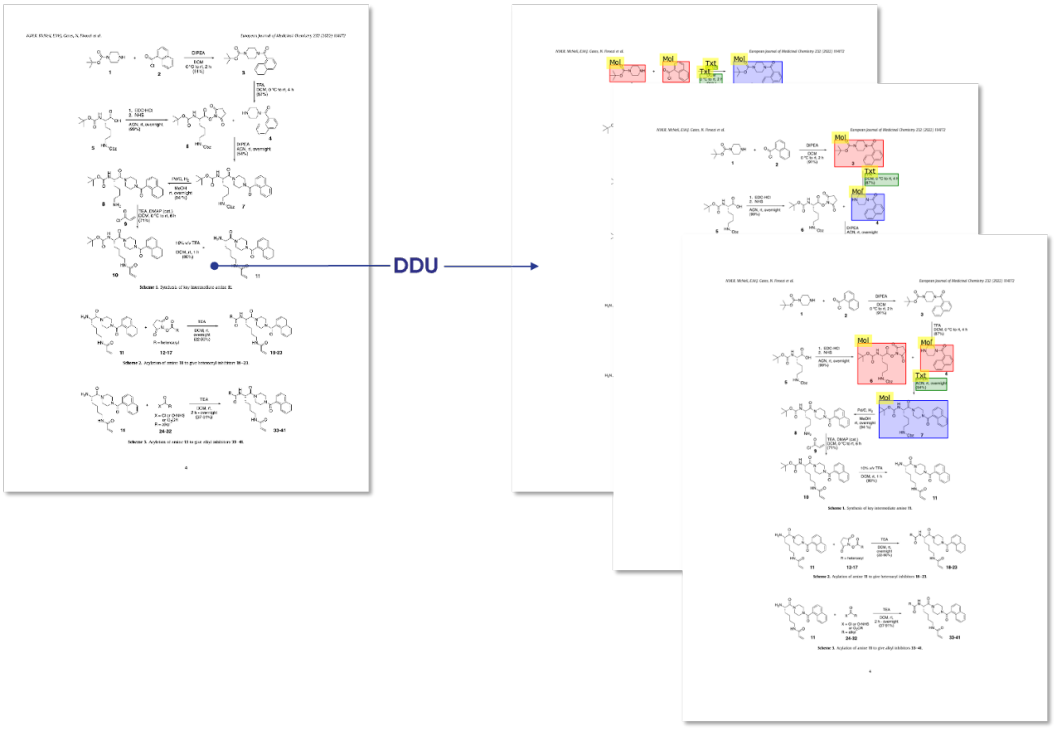


Figure 7: Expected Results/visualization

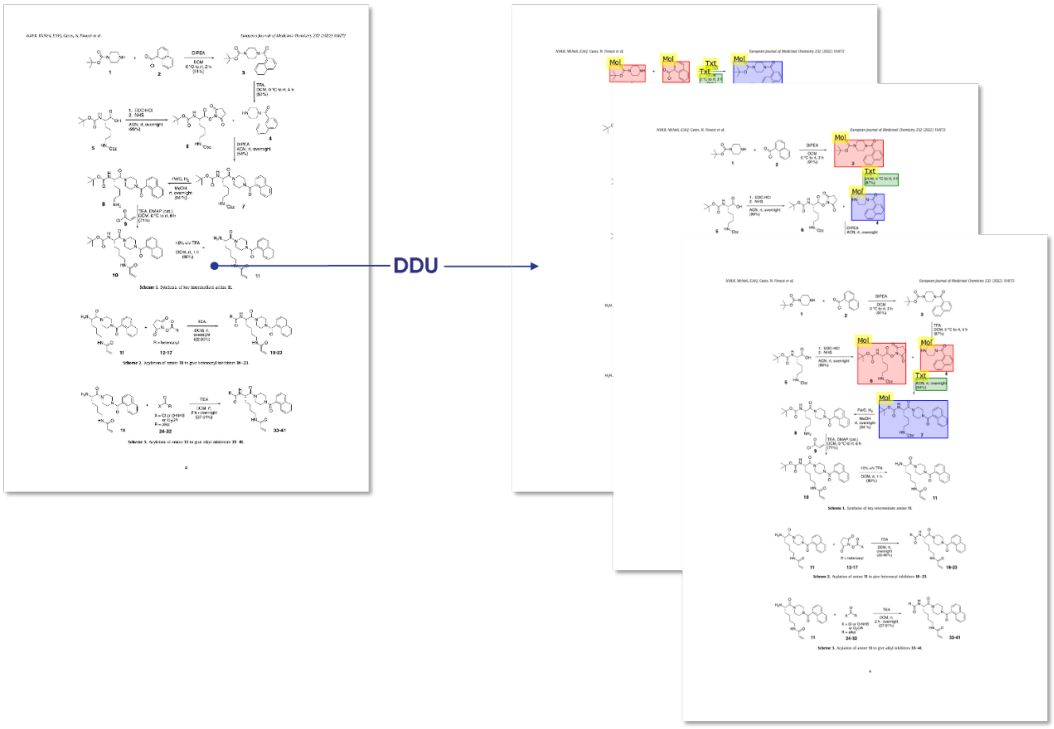


Figure 8: Expected Results/visualization

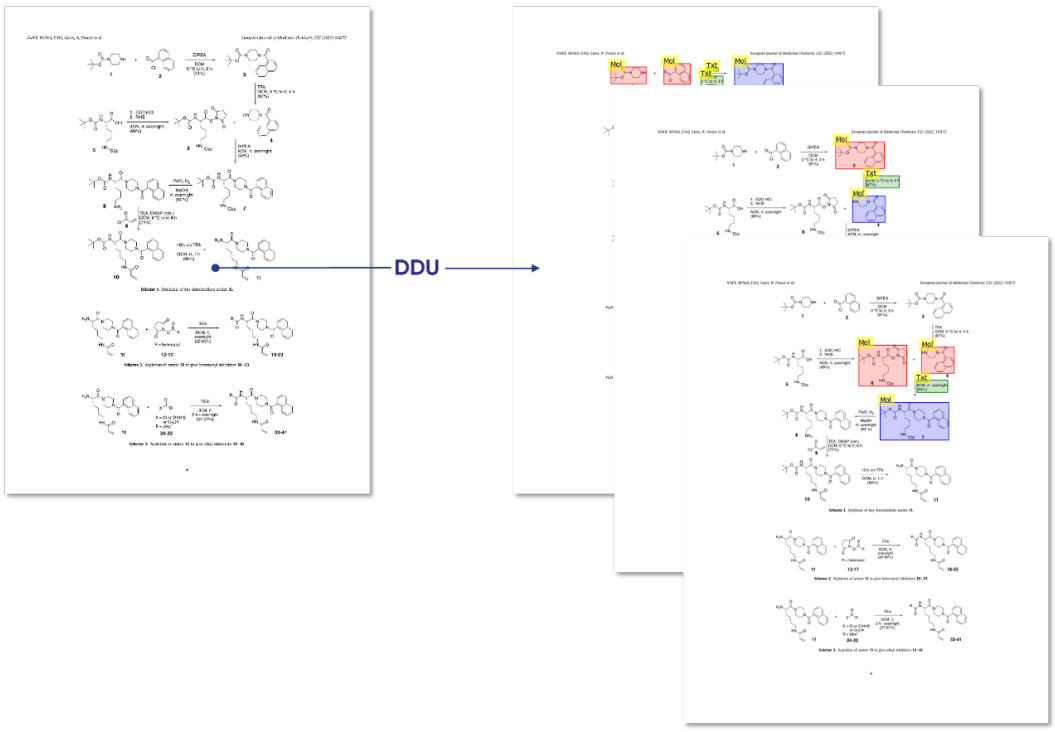


Figure 9: Expected Results/visualization

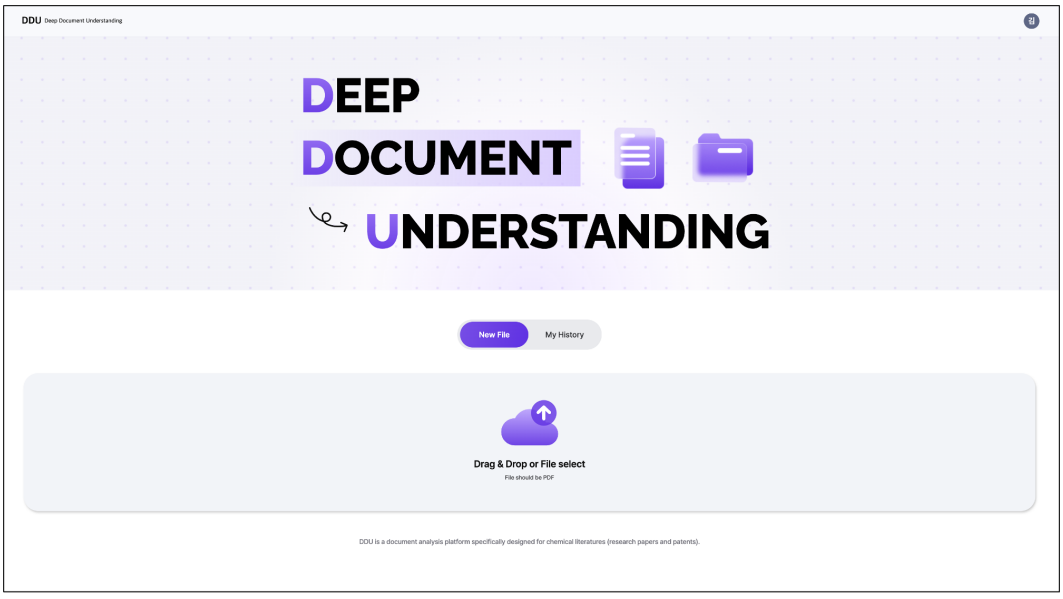


Figure 10: First Page to upload your PDF file.

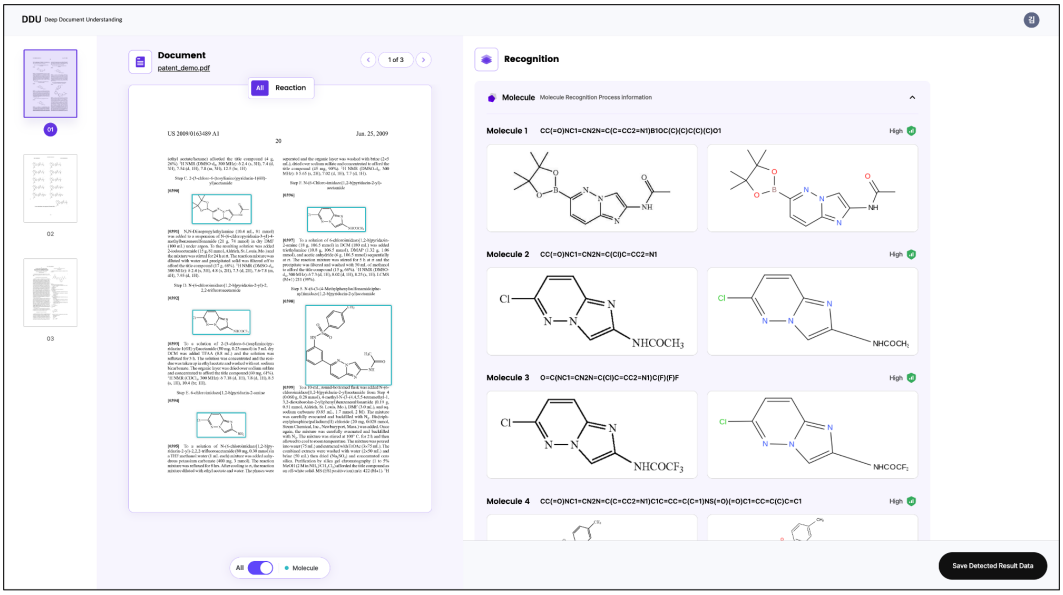


Figure 11: Processed results. On the left, you can check the ViDetect detection results of molecular regions. On the right, you can check the molecule conversion results in SMILES through ViMore with confidence score among (low, medium, high).

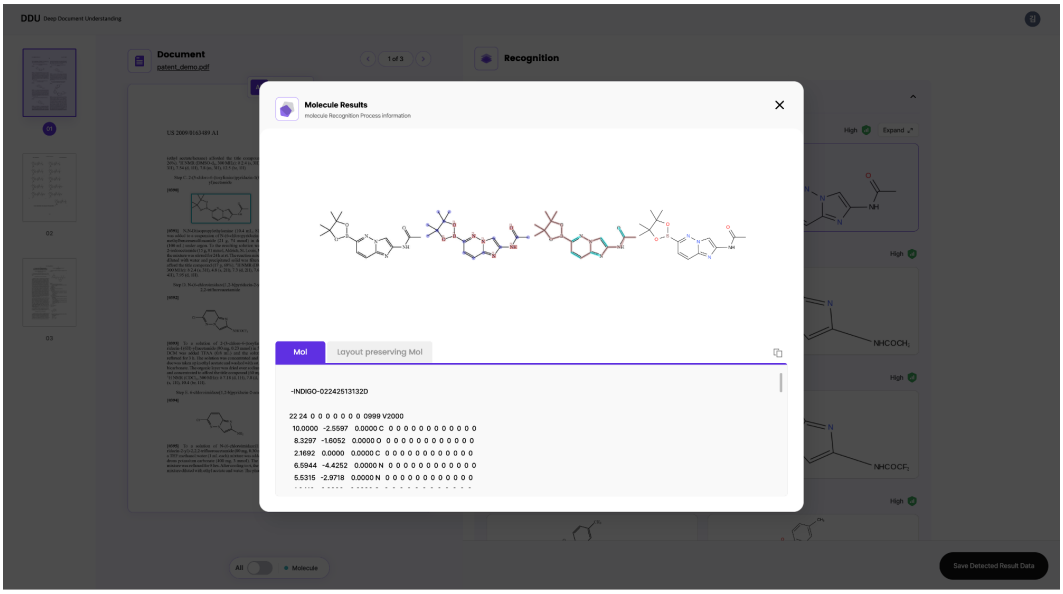


Figure 12: By clicking on expand, you can check the detailed model prediction results with Mol File.

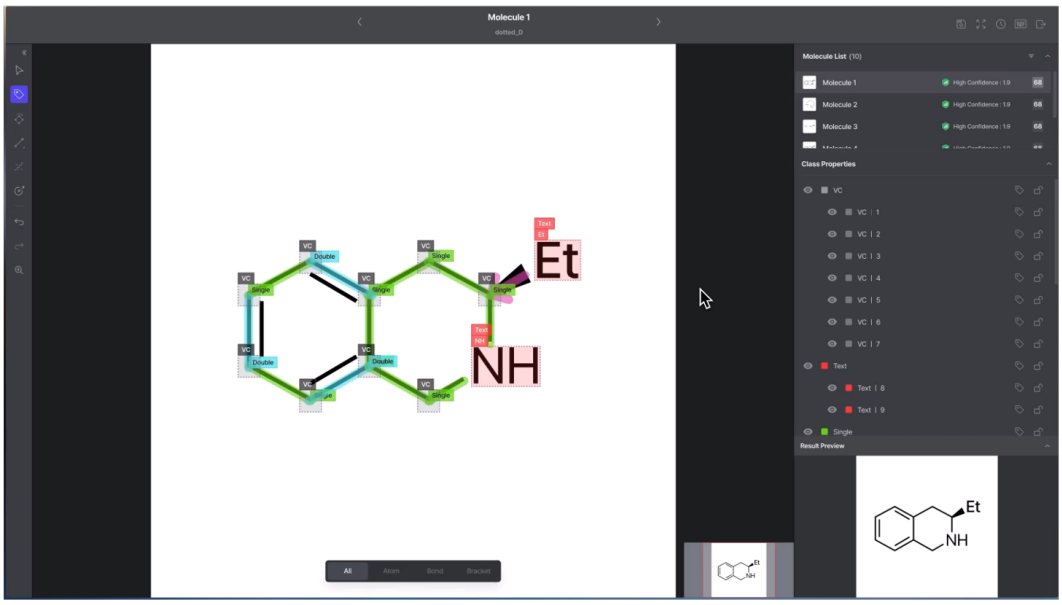


Figure 13: Additionally, you can activate the editing tool when there is a mistake in the prediction. By fixing the vision results, you can produce the correct mol file.

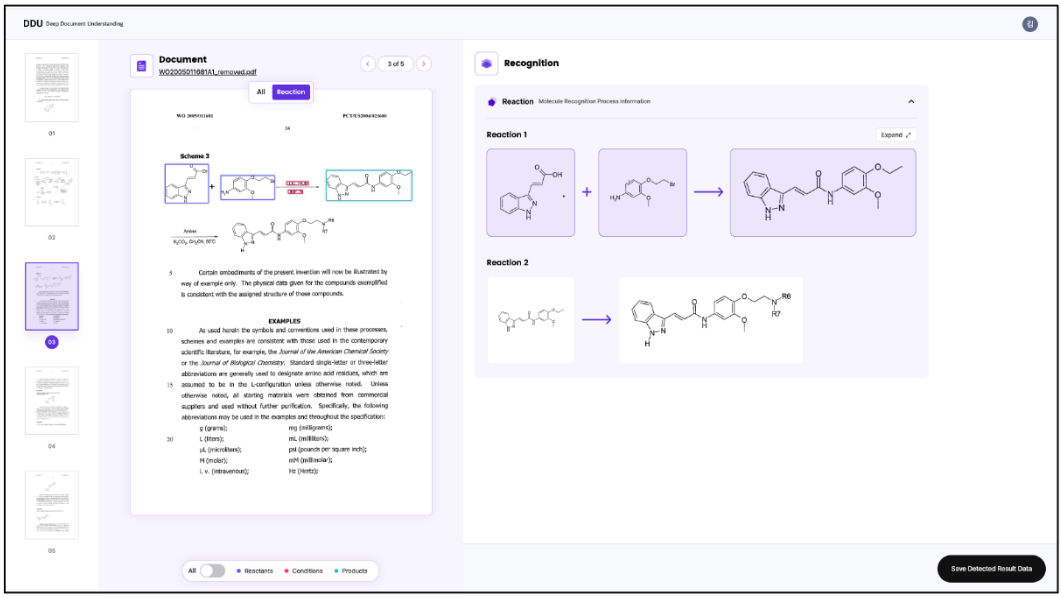


Figure 14: When there is reactions in the page, you can switch to Reaction mode. The detected reaction is visualized sequentially. The image is showing the first reaction detected in the page.

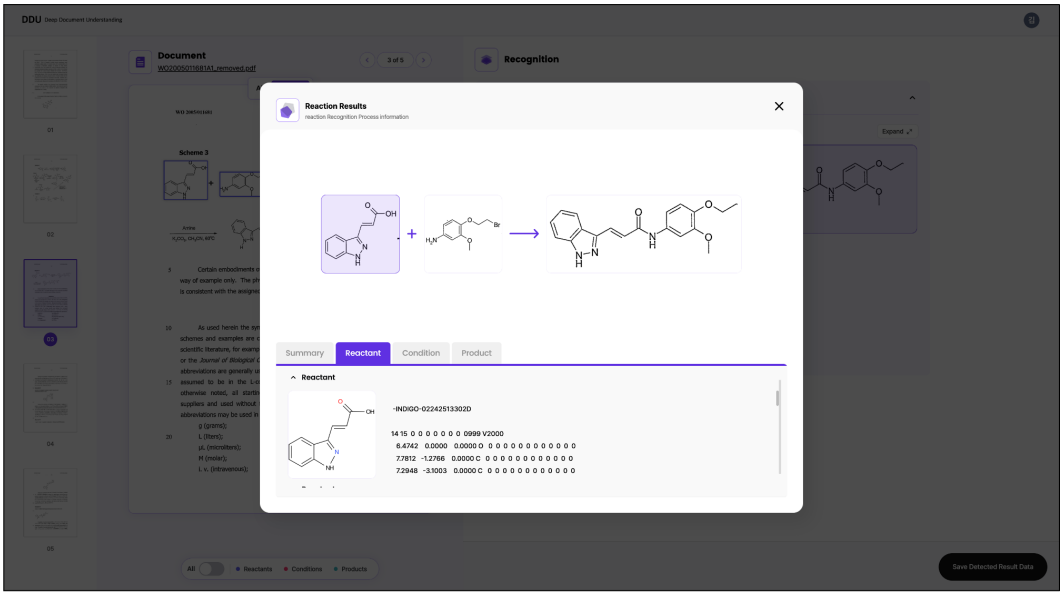


Figure 15: By clicking on expand, you can check the detected reactant, condition and product converted in Mol file format.

References

[1] A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. Gushurst, D. L. Grier, B. A. Leland, and J. Laufer. Description of several chemical structure file formats used by computer programs developed at molecular design limited. *Journal of chemical information and computer sciences*, 32(3):244–255, 1992.

[2] V. Fan, Y. Qian, A. Wang, A. Wang, C. W. Coley, and R. Barzilay. Openchemie: An information extraction toolkit for chemistry literature. *Journal of Chemical Information and Modeling*, 64 (14):5521–5534, 2024.

[3] L. Morin, M. Danelljan, M. I. Agea, A. Nassar, V. Weber, I. Meijer, P. Staar, and F. Yu. Molgrapher: graph-based visual recognition of chemical structures. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19552–19561, 2023.

[4] Y. Qian, J. Guo, Z. Tu, C. W. Coley, and R. Barzilay. Rxnscribe: a sequence generation model for reaction diagram parsing. *Journal of chemical information and modeling*, 63(13):4030–4041, 2023.

[5] Y. Qian, J. Guo, Z. Tu, Z. Li, C. W. Coley, and R. Barzilay. Molscribe: robust molecular structure recognition with image-to-graph generation. *Journal of Chemical Information and Modeling*, 63(7):1925–1934, 2023.

[6] K. Rajan, H. O. Brinkhaus, M. I. Agea, A. Zielesny, and C. Steinbeck. Decimer. ai: an open platform for automated optical chemical structure identification, segmentation and recognition in scientific publications. *Nature communications*, 14(1):5045, 2023.

[7] D. Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[8] D. M. Wilary and J. M. Cole. Reactiondataextractor 2.0: a deep learning approach for data extraction from chemical reaction schemes. *Journal of Chemical Information and Modeling*, 63 (19):6053–6067, 2023.