CEDe: Supplementary material

Rodrigo Hormazabal[1,2], Changyoung Park[1], Soonyoung Lee[1],
Sehui Han[1] Yeonsik Jo[1], Jaewan Lee[1], Ahra Jo[1]
Seunghwan Kim[1], Jaegul Choo[2], Moontae Lee[1], Honglak Lee[1]
[1]LG AI Research, [2]KAIST
{rodrigo, changyoung.park, soonyoung.lee, hansse.han
yeonsik.jo, jaewan.lee, ahra.jo
sh.kim, moontae.lee, honglak}@lgresearch.ai
jchoo@kaist.ac.kr

1 Datasheet for datasets [1]

1.1 Motivation

For what purpose was the dataset created? Was there a specific task in mind? *Was there a specific gap that needed to be filled? Please provide a description.* AI-based materials design is a rapidly growing area of research in the field of Chemistry. However, experimental data is scarce, while the obtention and indexation of data still constitute a major bottleneck in the materials' discovery process. Researchers mainly access information by extracting data from scientific documents, such as papers and patents.[2] Molecular images have been, and currently are, the preferred format for publishing discoveries and detailing structural information about new compounds. Previous approaches to automating the information extraction process from images utilized rule-based methods.[3] More recently, machine learning-based approaches have been explored for the same task.[4] However, until now, even state-of-the-art models have struggled to perform on par with traditional approaches due to being sample-inefficient. To overcome these issues, we present a collection of datasets, `CEDe` (**C**hemical **E**ntity **De**tection) in order to encourage research on more efficient molecular structure identification methods. These datasets aim to help train pipelines based on chemical instance recognition and subsequent graph reconstruction, showing higher sample efficiency over image-to-sequence translation models.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? The `CEDe` collection of datasets was developed by researchers at LG AI Research. Researchers with a background in chemistry, in conjunction with AI experts, worked to establish an annotation pipeline that included all the necessary information to train data-driven models for OCSR effectively..

Who funded the creation of the dataset? *If there is an associated grant, please provide the name of the grantor and the grant name and number.* This project was fully funded by LG AI Research. Researchers at LG AI Research generated these dataset annotations while conducting internal projects and decided to open-source their efforts in order to foster research in this particular area.

1.2 Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? *Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.* The `CEDe` datasets consist of molecular image-level metadata annotations, as well as bounding box annotations for each chemical entity contained within an image (i.e., nodes and edges). Each bounding box annotation

36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks.

also comes with their corresponding labels and necessary information for the molecular graph reconstruction.

How many instances are there in total (of each type, if appropriate)? The collection of datasets and the corresponding number of images are mentioned in the main paper. As for molecular entities, there is a total of 700,566 bounding box annotations with their corresponding labels/chemical information. Instance distributions for each dataset are also shown in the supplementary material, section 2.

Does the dataset contain all possible instances, or is it a sample (not necessarily random) of instances from a larger set? *If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).* All chemical entities appearing in the molecular images are labeled, and even random marks and signs that do not correspond to constituents of a molecular graph are annotated accordingly.

What data does each instance consist of? *"Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.* The image metadata and bounding box annotations are explained in the main paper; please refer to Fig. 3.

Is there a label or target associated with each instance? *If so, please provide a description.* All images are correspondingly labeled; please refer to Fig.3 in the main paper for a description.

Is any information missing from individual instances? *? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.* In order to reconstruct the underlying chemical graphs, no information is missing for individual instances.

Are there recommended data splits (e.g., training, development/validation, testing)? *? If so, please provide a description of these splits, explaining the rationale behind them.* We do not provide official recommendations for data splits in this paper.

Are there any errors, sources of noise, or redundancies in the dataset? *If so, please provide a description.* Bounding box annotations were performed by human expert annotators. This process was carried out with several inspections, and many efforts were taken to avoid errors as much as possible. However, errors may still exist, and we look to keep updating our datasets if missing annotations or mislabeled instances are found.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? *If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.* As explained in the main paper, the annotations we release correspond to chemical entity annotations appearing in existing open-source datasets; `UOB`, `USPTO`, `CLEF`, and `JPO`, which currently do not include this information. This labeling process is a one-time job that does not require continuous external source integration. Information about these external sources and the corresponding references are mentioned in the main paper.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals' non-public communications)? *If so, please provide a description.* The original molecular image datasets are open-sourced and extensively used by previous work. As far as we know, they do not contain any confidential data.

1.3 Collection Process

How was the data associated with each instance acquired? *Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.* The annotations were generated by recognizing the chemical information of atoms and bonds in each molecule image and subsequently transforming it to SMARTS fragments or pseudoatom text information.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? *How were these mechanisms or procedures validated?* The images used in this project were sourced from available open databases. Annotations were performed manually by experts with in-house annotation tools.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)? The authors of this paper, jointly with other members of LG AI Research, performed the data annotation process without any outsourcing.

Over what timeframe was the data collected? *Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)If not, please describe the timeframe in which the data associated with the instances was created.* The labeling process was performed over the course of 6 months as part of other related projects.

1.4 Dataset Uses

Has the dataset been used for any tasks already? *If so, please provide a description.* These dataset annotations have not been in any task outside LG AI Research.

Is there a repository that links to any or all papers or systems that use the dataset? *? If so, please provide a link or other access point.* Currently, there are no open-source projects that use the chemical entity annotations presented in this work.

What (other) tasks could the dataset be used for? As for now, this dataset can only be used for the task of recognizing molecular images. However, this data can be extended to work on a broader framework, for example linking chemical information appearing in the scientific literature in the form of images with related data in text form. In this case, we might use not only chemical entities' information appearing in molecular images but also other signs and markers that link to other parts of a document.

Are there tasks for which the dataset should not be used? *If so, please provide a description.* Not to our knowledge.

1.5 Dataset Distribution

Will the dataset be distributed to third parties outside the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. This data is provided free of charge, under an *"Attribution-Non Commercial 4.0 International (CC BY-NC 4.0)"* license, in order to foster research in OCSR-based applications.

How will the dataset be distributed (e.g., tarball on the website, API, GitHub)? *Does the dataset have a digital object identifier (DOI)?* We distribute the `CEDe` collection of datasets through an GCP Storage Bucket and, will open-source our baseline implementations in a GitHub repository https://github.com/rshormazabal/CEDe). Also, the dataset is registered in identifiers.org and has its corresponding prefix (lgai.cede).

When will the dataset be distributed? The plan to make `CEDe` publicly available after the corresponding NeurIPS review process of this work (before the start of NeurIPS 2022).

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? *If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions*
The annotations and corresponding labels will be distributed under an *"Attribution-Non Commercial 4.0 International (CC BY-NC 4.0)"* license, which allows for distribution, remix, and adaptation in any medium or format for noncommercial purposes only.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? *If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.* There are no imposed restrictions on the data released in this work.

1.6 Dataset Maintenance

Who will be supporting/hosting/maintaining the dataset? The dataset will be hosted and maintained by the authors of this work, researchers at LG AI Research.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)? Users can contact us through email or the related GitHub repository.

Is there an erratum? *If so, please provide a link or other access point.* We plan to release a website for data exploration, where users will be able to flag images/annotations.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? *If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?* We plan to keep the `CEDe` datasets updated for corrections, improvements, and extensions in the future. We plan to label other datasets that can help to recognize more complex chemical structures (polymers, reactions, etc.).

Will older versions of the dataset continue to be supported/hosted/maintained? *If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.* We will keep a log of updates in the GitHub repository and the soon-to-be-released website.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? *If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.* As mentioned before, we plan to release a website for data exploration, where users will be able to flag images/annotations. These tagged cases will be correspondingly fixed if necessary. However, currently, we are not considering any form of bigger scale extension or update method to our datasets from external users.

2 Dataset details

In the figures below, we present the chemical instance-level token distribution for the UOB, USPTO, CLEF and JPO datasets. SMARTS fragment instances like [C] or [O] are commonly present across all datasets and have a clear chemical meaning. However, pseudoatoms can have different interpretations depending on the context. Pseudoatoms can represent a cluster of atoms in a compressed format or work as markers for scientific document authors to point to a substructure within a chemical structure. Pseudoatoms representing atoms' clusters can permute in certain ways while keeping their meaning. For example, [CN] and [NC] represent the same group of atoms and connectivity. In the case of pseudoatoms used as markers, they tend not to permute and need to be identified as separate instance classes. The way we handle pseudoatoms annotations in this work is explained in the main manuscript (Sec. 3.1 - pseudoatoms). The number of molecular entity annotations per dataset is shown in table 1.

Table 1: Number of SMARTS fragment annotations and pseudoatom annotations for the CEDe. UOB dataset do not contain pseudoatom instances.

Datasets	SMARTS fragments	Pseudoatoms
UOB	188,311	0
USPTO	419,726	3,308
CLEF	63,897	362
JPO	24,766	196

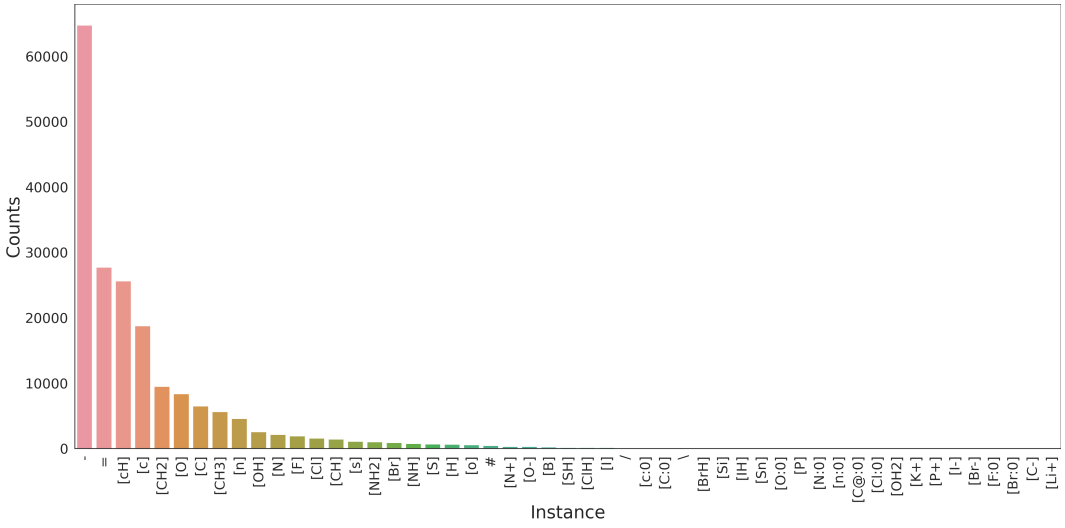

Figure 1: Instance class distribution for chemical entities that can be represented directly with SMARTS fragments for the **UOB dataset**. Does not consider pseudoatoms.

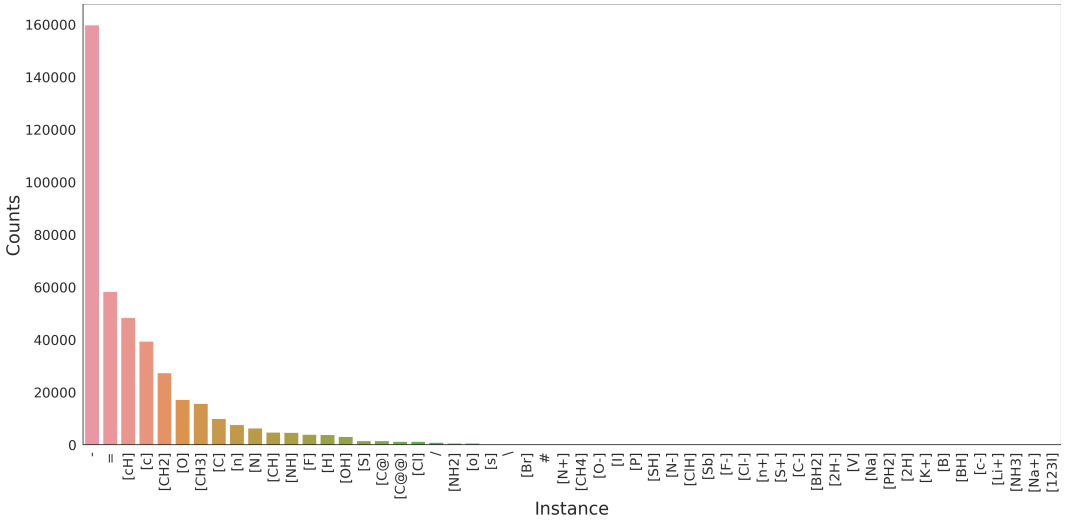

Figure 2: Instance class distribution for chemical entities that can be represented directly with SMARTS fragments for the **USPTO dataset**. Does not consider pseudoatoms.

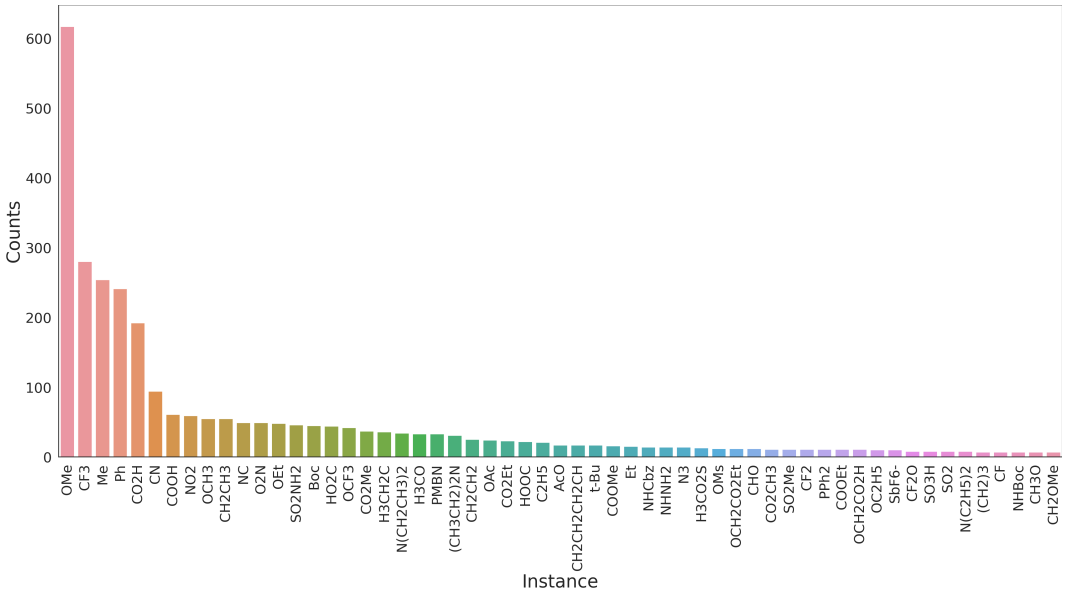

Figure 3: Instance class distribution for pseudoatoms in the **USPTO dataset**. Only shows the 60 most common instance classes from a total of 285.

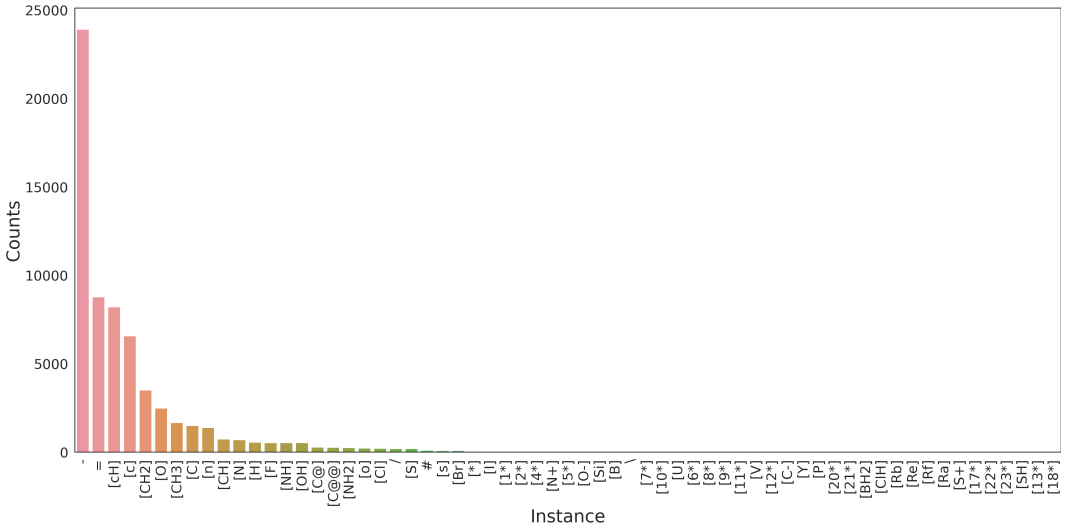

Figure 4: Instance class distribution for chemical entities that can be represented directly with SMARTS fragments for the **CLEF dataset**. Does not consider pseudoatoms.

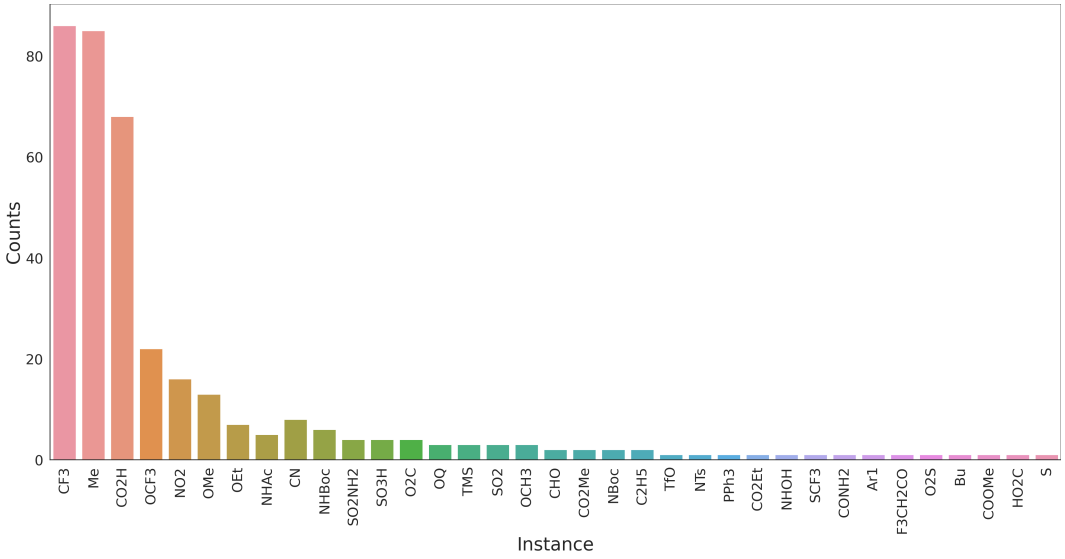

Figure 5: Instance distribution for pseudoatoms in the **CLEF dataset**. There are a total of 53 pseudoatom classes present in this dataset.

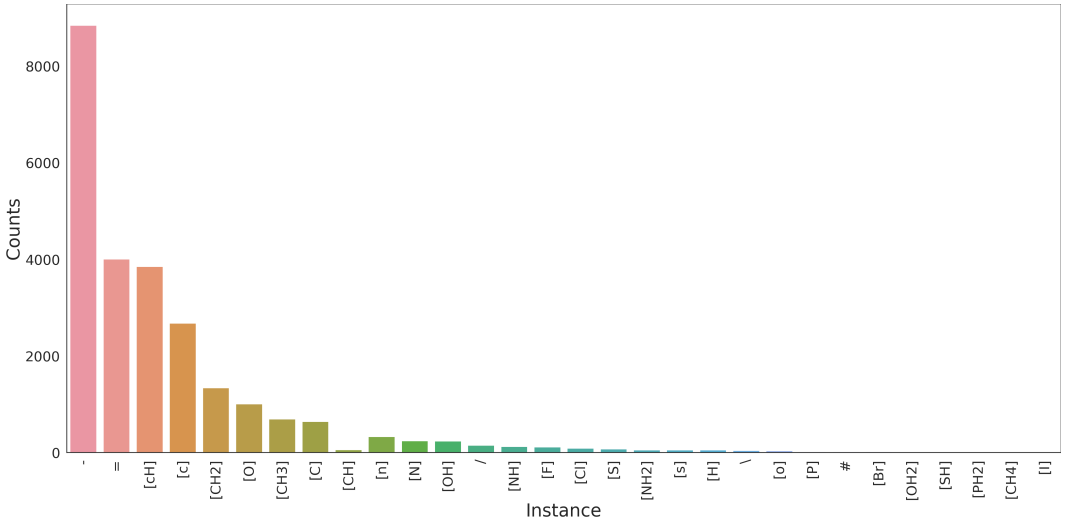

Figure 6: Instance class distribution for chemical entities that can be represented directly with SMARTS fragments for the **JPO dataset**. Does not consider pseudoatoms.

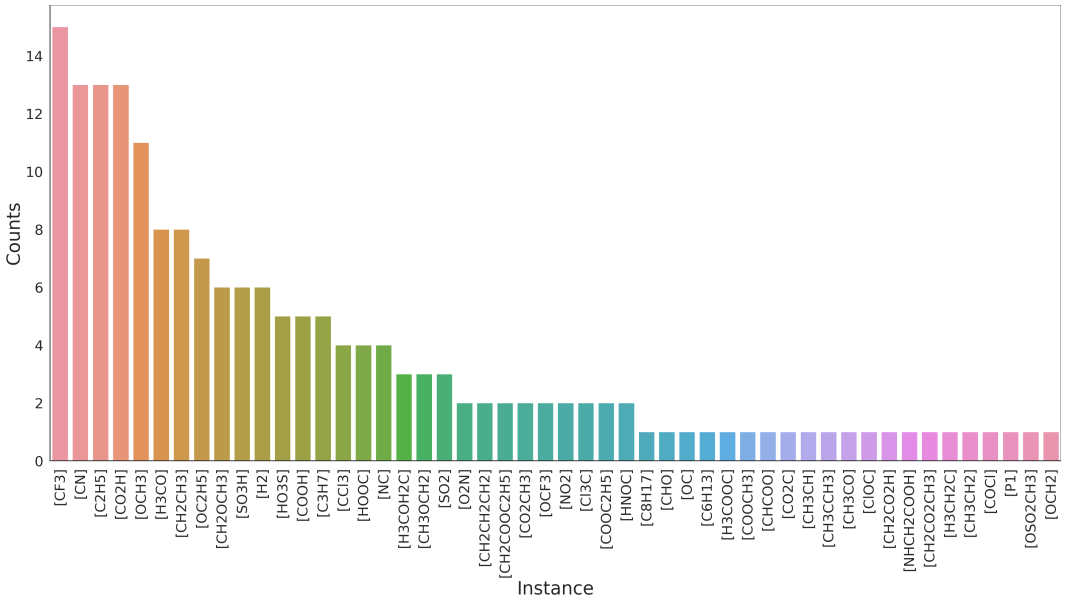

Figure 7: Instance class distribution for pseudoatoms in the **JPO dataset**. There are a total of 42 pseudoatom classes present in this dataset.

3 Detection models metrics: mean average precision per class

Mean average precision (mAP) for detection models presented in section 4 of this work (Main paper, Table 2) is shown in the table below. In addition, a comparison between experiments using synthetic data for pretraining (10K) and models from random initialization are presented. The column *'Diff'* shows the difference in percentage with respect to models without pre-training.

Table 2: Mean average precision per class (mAP) for detection models used in this work. **"No pre-trained"** uses a subset of CEDe to train (2575 images) and **"Pre-trained"** uses 10K synthetic images for pretraining and finetunes with the same subset of CEDe as "no pre-trained".

Instance class	RCNN (mAP)			DETR (mAP)		
	No pre-trained	Pre-trained	**Diff**	No pre-trained	Pre-trained	**Diff**
[STEREOE]	37.33	53.05	**+42.12%**	32.51	47.63	**+46.52%**
[C@]	34.74	48.33	**+39.12%**	34.52	49.36	**+42.98%**
[C@@H]	42.42	57.76	**+36.15%**	39.44	52.12	**+32.15%**
down	46.69	61.92	**+32.61%**	44.89	60.67	**+35.18%**
[C@H]	52.14	67.32	**+29.12%**	49.26	59.67	**+21.14%**
[o_arom]	50.25	63.53	**+26.43%**	49.48	63.78	**+28.88%**
up	62.96	77.70	**+23.41%**	61.55	73.33	**+19.15%**
[C]	66.79	82.31	**+23.23%**	65.58	82.94	**+26.46%**
[STEREOZ]	45.30	55.46	**+22.42%**	42.31	52.45	**+23.98%**
[Me]	69.28	82.07	**+18.47%**	68.00	77.41	**+13.85%**
other_signs	31.23	36.63	**+17.28%**	28.22	35.23	**+24.86%**
#	73.01	84.41	**+15.61%**	69.84	81.89	**+17.26%**
[CO2H]	68.01	78.57	**+15.52%**	62.53	75.01	**+19.96%**
[CH]	63.34	73.10	**+15.42%**	61.18	70.53	**+15.29%**
[bond_arom]	64.80	73.60	**+13.58%**	61.22	67.52	**+10.30%**
[CH2]	69.08	78.12	**+13.09%**	67.53	80.96	**+19.89%**
[n_arom]	70.41	79.50	**+12.91%**	69.12	80.14	**+15.93%**
either	48.51	54.66	**+12.66%**	47.80	56.47	**+18.13%**
[CN]	59.12	66.60	**+12.65%**	57.68	63.41	**+9.92%**
[NH]	74.59	83.19	**+11.52%**	71.86	82.59	**+14.93%**
[H]	64.91	71.31	**+9.86%**	62.95	69.39	**+10.24%**
[F]	71.72	78.47	**+9.41%**	71.36	79.38	**+11.23%**
[OMe]	76.91	82.91	**+7.80%**	74.04	87.31	**+17.93%**
-	75.47	80.87	**+7.16%**	69.47	77.26	**+11.20%**
[s_arom]	68.19	72.87	**+6.86%**	66.51	71.52	**+7.53%**
[N]	74.28	79.06	**+6.43%**	72.82	75.11	**+3.15%**
[Cl]	74.30	78.93	**+6.23%**	73.18	80.45	**+9.94%**
[CH3]	66.93	71.07	**+6.19%**	60.47	64.41	**+6.53%**
[cH_arom]	70.87	75.21	**+6.13%**	68.27	74.22	**+8.72%**
[OH]	74.35	78.90	**+6.12%**	70.22	72.50	**+3.25%**
[c_arom]	70.23	74.50	**+6.09%**	63.90	67.18	**+5.13%**
[CF3]	66.02	69.48	**+5.25%**	63.24	66.07	**+4.48%**
[NH2]	75.66	79.07	**+4.51%**	70.16	75.08	**+7.00%**
[Br]	72.87	75.75	**+3.95%**	65.92	67.85	**+2.93%**
[Ph]	74.27	77.15	**+3.87%**	71.97	73.31	**+1.87%**
[nH_arom]	61.93	64.05	**+3.42%**	56.70	58.14	**+2.54%**
[NO2]	73.61	76.06	**+3.33%**	71.13	74.63	**+4.92%**
[O]	74.12	76.37	**+3.03%**	72.85	73.56	**+0.98%**
[S]	69.94	71.97	**+2.90%**	66.75	67.61	**+1.29%**
other	50.91	52.09	**+2.31%**	49.61	50.10	**+0.99%**
=	76.91	78.53	**+2.11%**	70.70	72.36	**+2.34%**
[I]	66.97	68.21	**+1.86%**	63.79	65.92	**+3.34%**
Average mAP	63.84	71.21	**+13.05%**	60.96	68.49	**+13.90%**

4 Generation augmentations effect

The effect of augmentation in the synthetic data generation process was explored. For every augmentation, an experiment consisting of data generation, pre-training, and finetuning was done. When certain random augmentations are not used, they are set to the mean value across the original sample range. `Pseudatom ratio` refers to the percentage of compounds where an atom is replaced by a randomly sampled superatom/pseudoatom instance, which shows the most significant impact on the performance over real data. `Rotation` and `XY sheer` are set before image file is generated, so they do not perturb the letters orientation (contrary to what happens when applying rotation augmentations at training time).

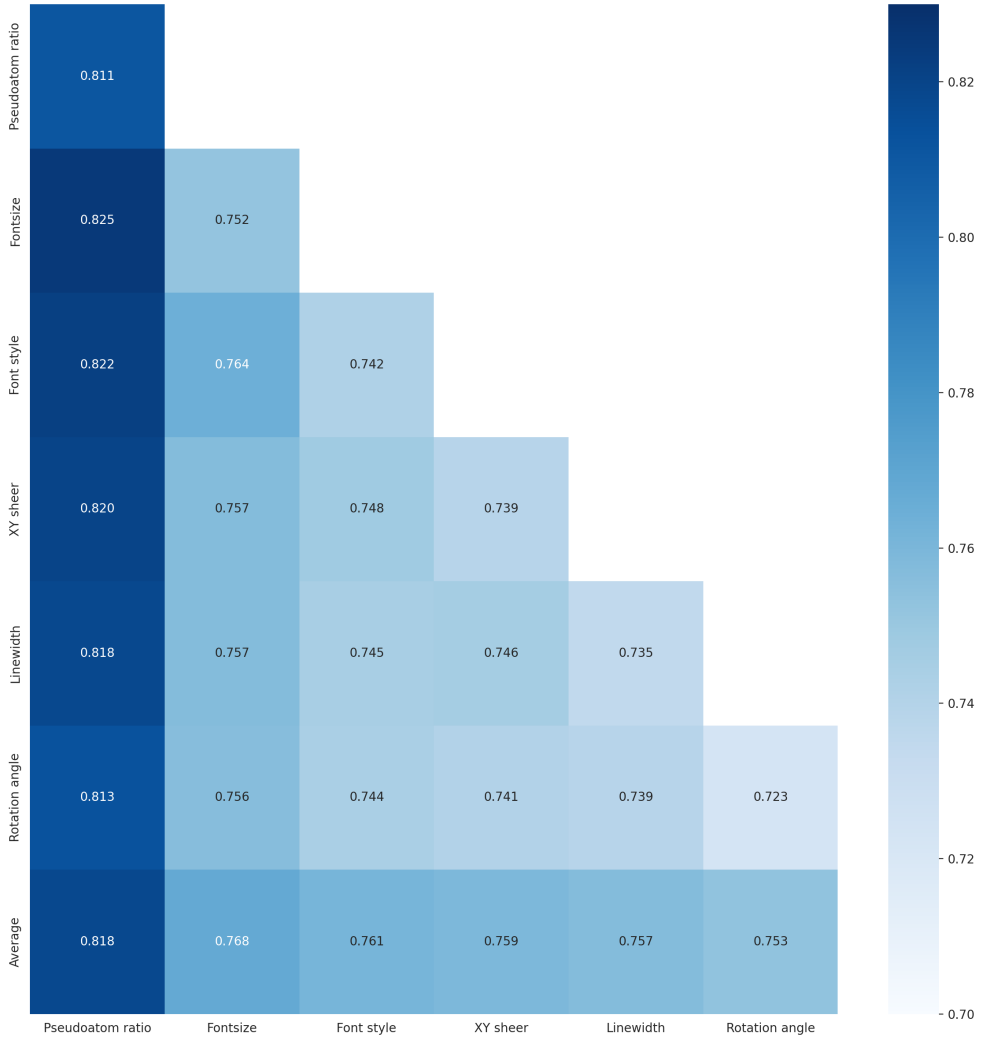

Figure 8: Effect of style-augmentations for synthetic data in fine-tuning tasks' performance is shown. Each augmentation by itself (diagonal) and, pairs of augmentations are shown. Also, the average performance of each augmentation while paired with others is presented.

5 Previous work baselines

Performance of previous work [4, 5, 6] in the test splits utilized for our baseline experiments is shown in Table 3. Baselines presented in our work are vanilla implementations of these approaches that do not rely on any pre-trained model for image featurization or string representation encoding. As an example, in previous works, such as Img2Mol and DECIMER, pre-trained models are used to

generate image features as a preprocessing step (Inception V3[7]), decoding SMILES sequences (CDDD[8]) and for encoding SMILES strings into representations (DeepSMILES[9], SELFIES[10]) are used, which adds a lot of complexity to the benchmarking scheme. Here, we present the accuracy of these models calculated with their open-source implementations and pre-trained models. The amount of data used for training each of these models is shown in the column names.

Table 3: Accuracy of open-source implementations of previous work for the OCSR task. Models were tested against the splits used for our main experiments. The amount of data used to train each model is also shown.

	img2mol@11M	**Chemgrapher@140K**	**DECIMER@15M**
UOB	0.612	**0.832**	0.538
USPTO	0.459	**0.809**	0.374
CLEF	0.441	**0.755**	0.260
JPO	0.386	**0.533**	0.272

6 Baselines implementation details

Models We present six baselines in our work in order to show how frameworks based on chemical entity detection followed by graph reconstruction outperform image-to-sequence methods, even with orders of magnitude fewer data. Architecture details corresponding to these baselines are presented below. Also, we plan to open-source a PyTorch [11] implementation of these baselines and provide trained model weights to reproduce our results. Details about the corresponding models, optimizers, and hyperparameters will be provided as config files at `https://github.com/rshormazabal/CEDe`. In addition, the sampled images for the fine-tuning task and the corresponding sampling procedure will also be available.

Image-to-SMILES For the image-to-SMILES models, we use as image feature structures CNN backbones, specifically ResNet101[12]. As for the autoregressive generation module, we use vanilla PyTorch GRU[13], attention [14] and transformer implementations [15]. For the tokenization scheme, we follow previous work [6]. Details will be available in the aforementioned repository.

Chemical entity detection For the detection modules (DETR [16] and Faster-RCNN [17]), we based our implementations on the Detectron2 vision library [18]. For the connectivity prediction task, we also use a vanilla PyTorch implementation of a transformer. We calculate pairwise interactions between the bounding box representations generated by the transformer in order to predict connected instances. Details about this pipeline and the rule-based connectivity prediction process will be available in the GitHub repository.

References

[1] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

[2] Zach Jensen, Soonhyoung Kwon, Daniel Schwalbe-Koda, Cecilia Paris, Rafael Gómez-Bombarelli, Yuriy Román-Leshkov, Avelino Corma, Manuel Moliner, and Elsa A Olivetti. Discovering Relationships between OSDAs and Zeolites through Data Mining and Generative Neural Networks. *ACS Central Science*, 7(5):858–867, 2021.

[3] Igor V Filippov and Marc C Nicklaus. Optical structure recognition software to recover chemical information: Osra, an open source solution, 2009.

[4] Martijn Oldenhof, Adam Arany, Yves Moreau, and Jaak Simm. Chemgrapher: optical graph recognition of chemical compounds by deep learning. *Journal of Chemical Information and Modeling*, 60(10):4506–4517, 2020.

[5] Djork-Arné Clevert, Tuan Le, Robin Winter, and Floriane Montanari. Img2mol-accurate smiles recognition from molecular graphical depictions. 2021.

[6] Kohulan Rajan, Achim Zielesny, and Christoph Steinbeck. Decimer: towards deep learning for chemical image recognition. *Journal of Cheminformatics*, 12(1):1–9, 2020.

[7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[8] Robin Winter, Floriane Montanari, Frank Noé, and Djork-Arné Clevert. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chemical science*, 10(6):1692–1701, 2019.

[9] Noel O'Boyle and Andrew Dalke. Deepsmiles: an adaptation of smiles for use in machine-learning of chemical structures. 2018.

[10] Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024, 2020.

[11] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[13] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[14] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.

[15] Simao Herdade, Armin Kappeler, Kofi Boakye, and Joao Soares. Image captioning: Transforming objects into words. *Advances in Neural Information Processing Systems*, 32, 2019.

[16] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[18] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.