SPATIAL INFORMATION INFERENCE IMPROVES ACTIVATION ENERGY PREDICTIONS

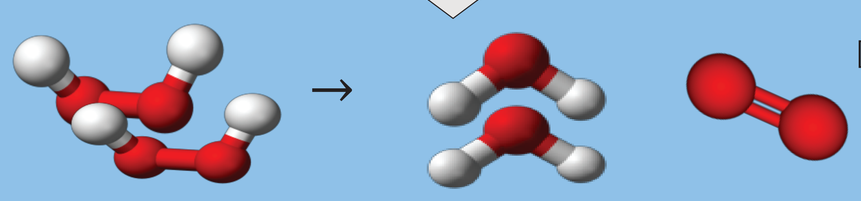
Introduction

- Activation energy is an important property in chemistry
- Expensive to conduct experiment
- Previous models predict using only chemical equation
- Want to improve accuracy by implicitly using spatial information[1]
- Work done on molecule properties, not reactions[2]

Methodology

Multi-step training process
Generate spatial information, processing the 2D reaction into 3D, using *ETKDG* (RDKit) to generate conformers[3]

$$2H_2O_2 \rightarrow 2H_2O + O_2$$



[5]

Figure 1: Example spatial information generated

The data was fed into two models as such:

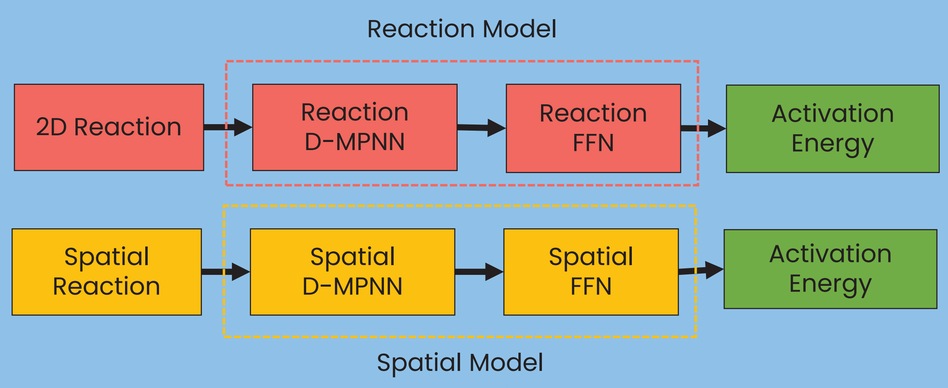


Figure 2: Operation of the two models.

Extract the *D-MPNN* from the two models,
Train to be similar with contrastive learning
Set pairs from same reactions to zero loss (NTXloss)
Set pairs from different reactions to high loss

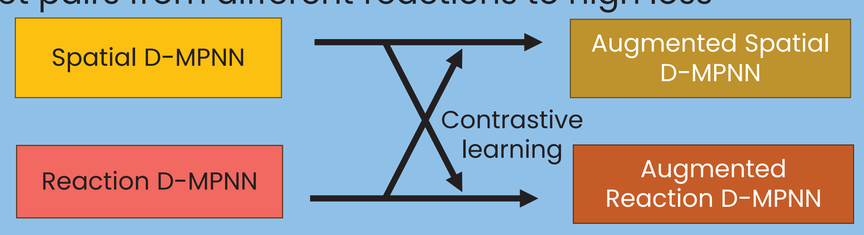


Figure 3: Creation of the augmented encoders.

Creates augmented (contrastively-trained) *D-MPNN*s which can implicitly encodes spatial information
Use augmented reaction *D-MPNN*

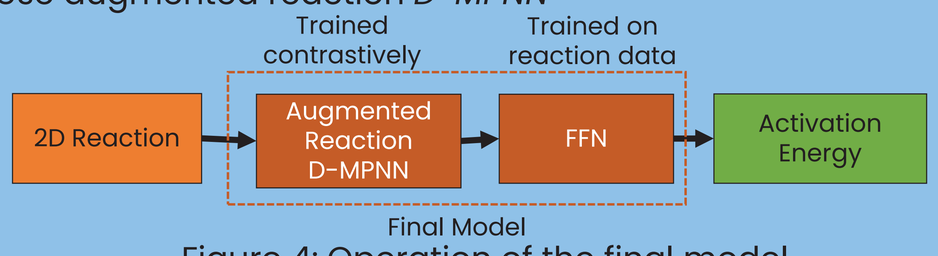


Figure 4: Operation of the final model.

Results and Discussion

Our model uses the *ωB97X-D3* dataset[4].

	Final Model	Reaction Model
MAE	6.02	6.12
RMSE	9.46	9.56

Figure 5: Comparison between final and reaction model

Paired-t test on the hypothesis that the final model gives less absolute error yielded p=0.002

More accurate Activation Energy Predictions via implicit accounting of spatial information

Generating spatial information on test set took 3 minutes
Generating predictions took 3 seconds
60-fold improvement

Spatial Information Inference gives quick Activation Energy Predictions

Future Work

- Include higher quality spatial information
- Include bond angle as a parameter
- Experiment on datasets of larger molecules

Applications

- Picking candidate pathways for organic synthesis
- E.g. Drug manufacturing or novel materials
- If reaction has too high activation energy, should consider alternative reaction

Conclusion

- Created a model that can predict activation energy
- Improves accuracy with spatial information inference (p=0.002)
- Maintaining same high prediction speed

References

[1] Heid, Esther, and William H. Green. "Machine learning of reaction properties via learned representations of the condensed graph of reaction." Journal of Chemical Information and Modeling 62.9 (2021): 2101-2110
[2]Stärk, Hannes, et al. "3d infomax improves gnns for molecular property prediction." International Conference on Machine Learning. PMLR, 2022.
[3]Conformer Generation Using RDKit, 2012. www.rdkit.org/UGM/2012/Ebejer_20110926_RDKit_1stUGM.pdf. (accessed 12/2023)
[4] Grambow, Colin A., Lagnajit Pattanaik, and William H. Green. "Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry." Scientific data 7.1 (2020): 137.
[5]Ball and stick model of molecules in decomposition of hydrogen peroxide: https://en.wikipedia.org/wiki/Hydrogen_peroxide#/media/File:Hydrogen-peroxide-3D-balls.png https://en.wikipedia.org/wiki/File:Water-3D-balls-A.png https://commons.wikimedia.org/wiki/File:Ossigeno_molecolare.jpg (all accessed 02/2024)

Members:
Chua Wee Chong, NUS High School of Mathematics and Science
Liu Wenkai, NUS High School of Mathematics and Science
Mentors:
Alvin Liew, DSO National Laboratories
Dr Chieu Hai Leong, DSO National Laboratories

YDSP Young Defence Scientists Programme
DSTA Defence Science & Technology Agency
DSO NATIONAL LABORATORIES