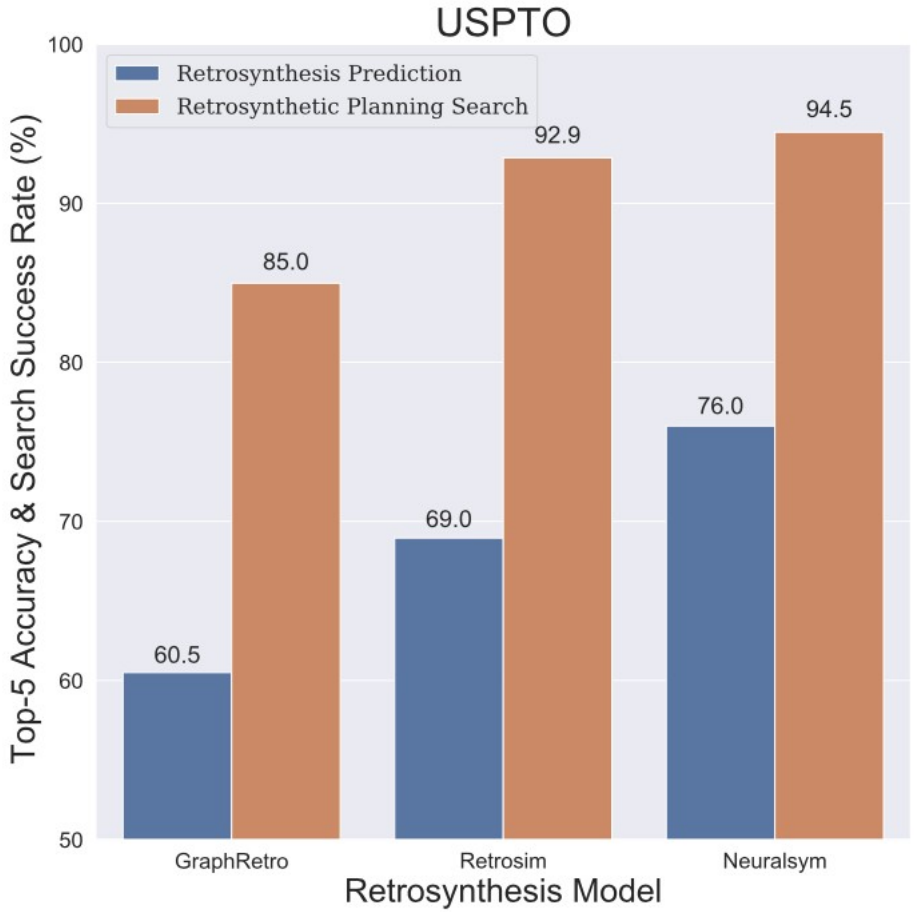
FusionRetro: Molecule Representation Fusion via In-Context Learning for Retrosynthetic Planning

Songtao Liu [1] **Zhengkai Tu** [*2] **Minkai Xu** [*3] **Zuobai Zhang** [*4 5]
Lu Lin [1] **Rex Ying** [6] **Jian Tang** [4 7 8] **Peilin Zhao** [9] **Dinghao Wu** [1]

1. Introduction

- Retrosynthesis.
- Current methods and Evaluation metrics.
- Novelty – Exact Set-wise match!
- Also introduce a dataset (58099 routes) + metric.
- Beats SOTA.

- (Contribution) New Model architecture.
- (Contribution) Many new experiments.

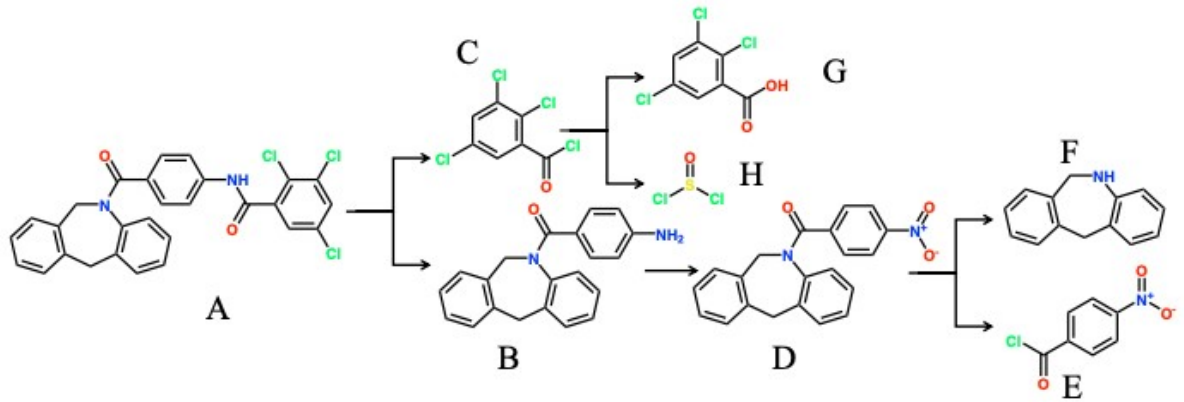


2. Related Work

- Single Step retrosynthesis (Template based vs Template free vs Semi Template based),
- Search Algorithms in Retrosynthetic Planning,
- Evaluation (efficiency vs quality).

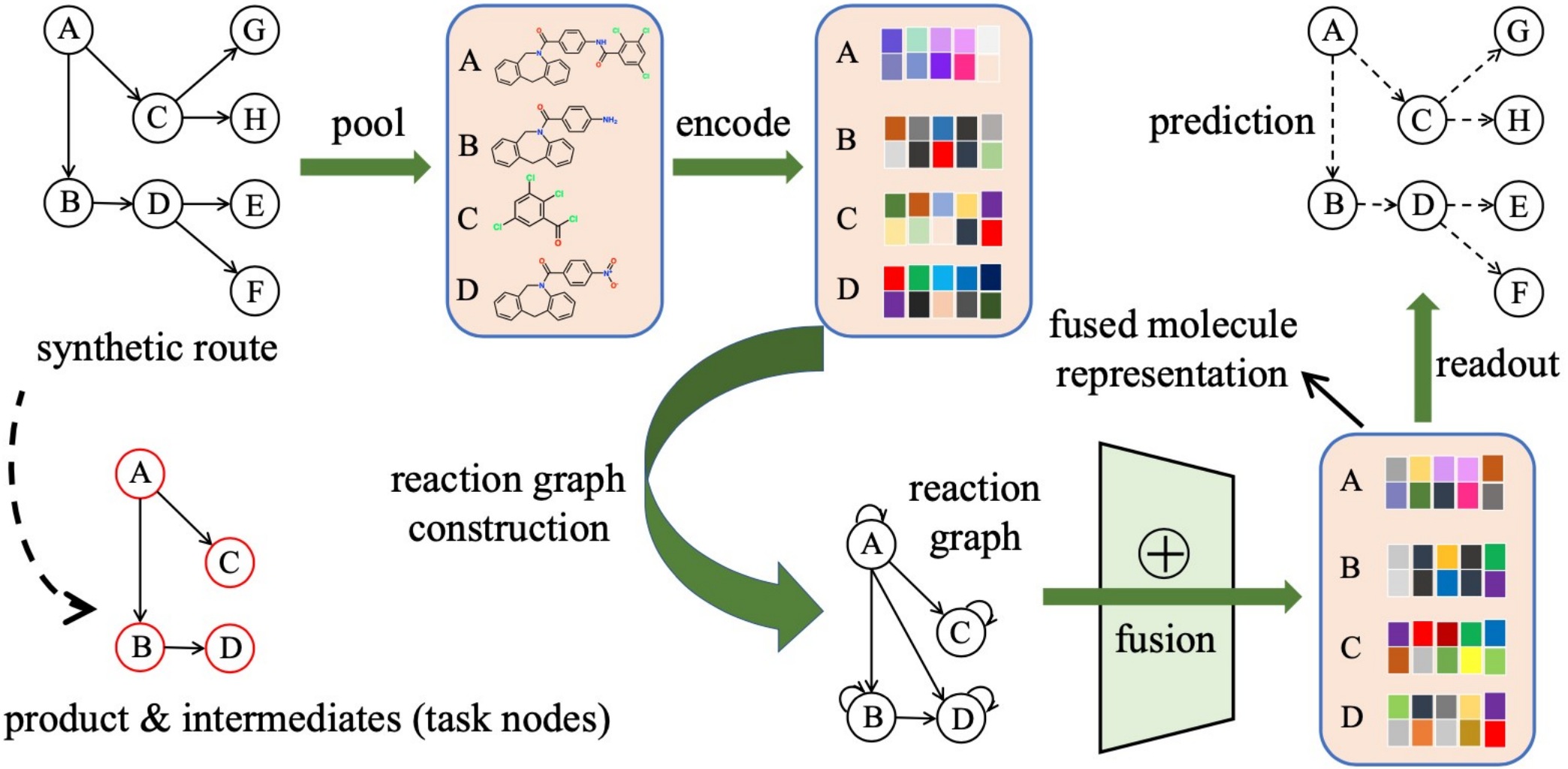
3. Notation

- Starting material – $\mathcal{M}$ (commercially purchasable molecules, ZINC),
- Synthetic Route – $G = \{T, R, I, \tau\}$ "reaction graph",

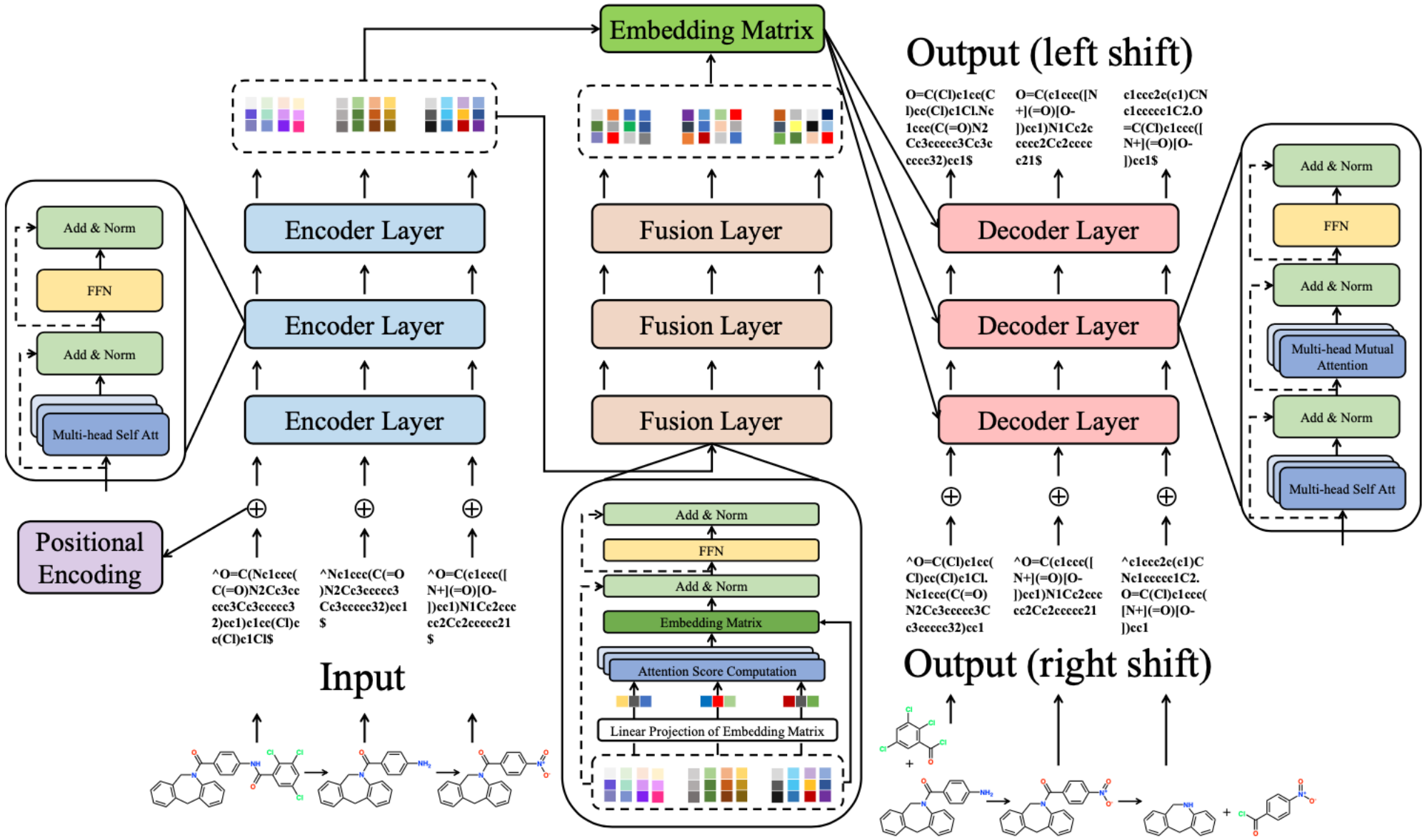


- Single-step retrosynthesis, $T \rightarrow R$,
- Retrosynthetic planning, $T \rightarrow I \rightarrow R$.

4. FusionRetro.



4. FusionRetro.



4. FusionRetro

- Loss –

$$\mathcal{L}(y, p) = -\sum_{i=1}^{n}\sum_{j=1}^{K} y_{ij} \log\left(p_{ij}\right)$$

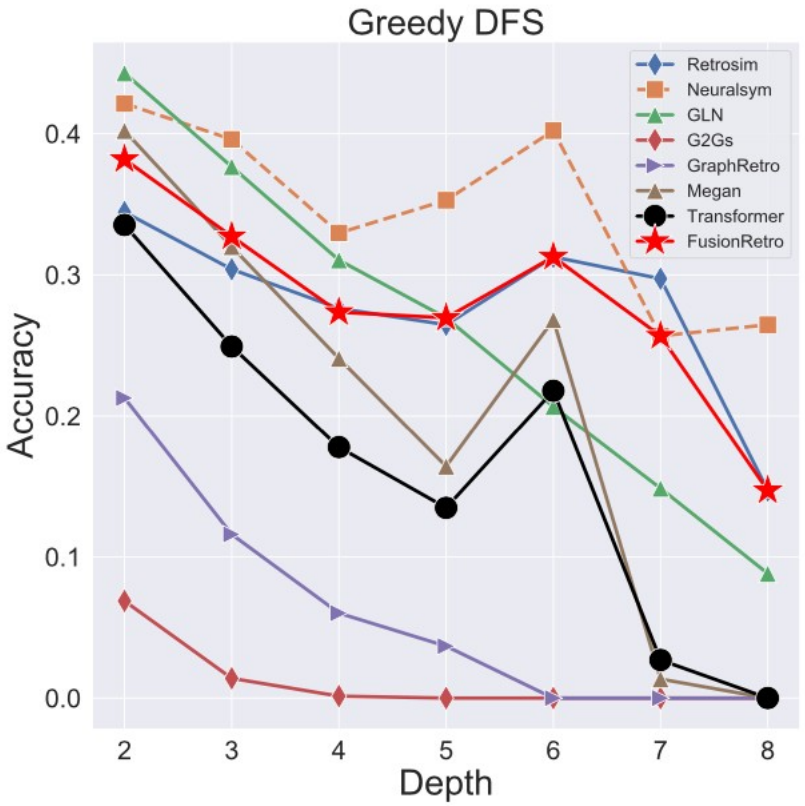
- Inference –

Algorithm 1 Inference given a target molecule

Input: Target molecule T, starting material set $\mathcal{S}$
Initialize reactant set $\mathcal{R} = \{\}$, path set $\mathcal{L} = \{\}$
Put the initial path $[T]$ into $\mathcal{L}$
while $\mathcal{L}$ is not a empty set **do**
  Take an path l from $\mathcal{L}$
  Predict the reactants r_l for expansion given l
  for reactant $r_l^{(i)}$ in r_l **do**
    if $r_l^{(i)} \in \mathcal{S}$ **then**
      Put $r_l^{(i)}$ into $\mathcal{R}$
    else
      Generate a new path $l' = l + [r_l^{(i)}]$
      Put l' into $\mathcal{L}$
    end if
  end for
end while
return predicted reactant set $\mathcal{R}$

5. Experiments

Search Algorithm	Retro*					Retro*-0					Greedy DFS
	Top-1	Top-2	Top-3	Top-4	Top-5	Top-1	Top-2	Top-3	Top-4	Top-5	Top-1
Template-based											
Retrosim (Coley et al., 2017)	35.1	40.5	42.9	44.0	44.6	35.0	40.5	43.0	44.1	44.6	31.5
Neuralsym (Segler & Waller, 2017)	**41.7**	**49.2**	52.1	53.6	54.4	**42.0**	**49.3**	52.0	53.6	54.3	**39.2**
GLN (Dai et al., 2019)	39.6	48.9	**52.7**	**54.6**	**55.7**	39.5	48.7	**52.6**	**54.5**	**55.6**	38.0
Template-free											
G2Gs (Shi et al., 2020)	5.4	8.3	9.9	10.9	11.7	4.2	6.5	7.6	8.3	8.9	3.8
GraphRetro (Somnath et al., 2021)	15.3	19.5	21.0	21.9	22.4	15.3	19.5	21.0	21.9	22.2	14.4
Megan (Sacha et al., 2021)	18.8	29.7	37.2	42.6	45.9	19.5	28.0	33.2	36.4	38.5	32.9
Transformer (Karpov et al., 2019)	31.3	40.4	44.7	47.2	48.9	31.2	40.5	45.1	47.3	48.7	26.7
FusionRetro	**37.5**	**45.0**	**48.2**	**50.0**	**50.9**	**37.5**	**45.0**	**48.3**	**50.2**	**51.2**	**33.8**

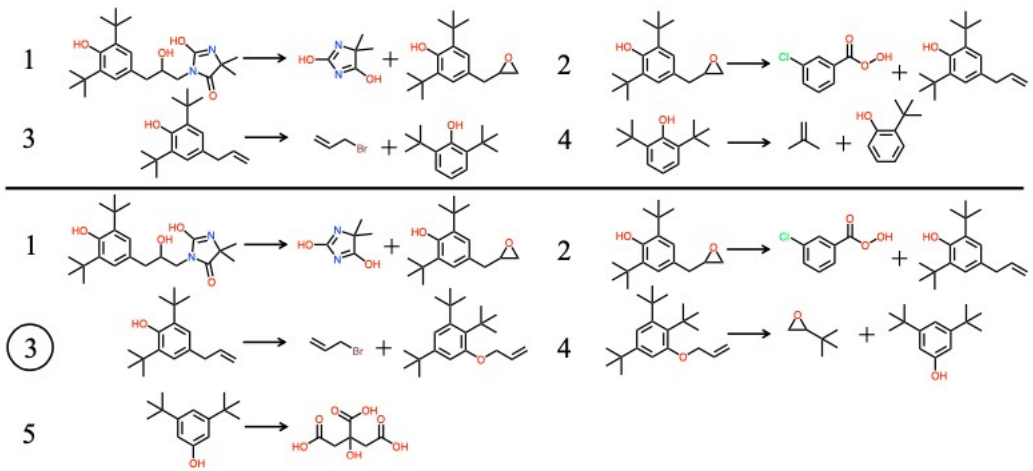


5. Experiments

Table 2. Summary of retrosynthesis prediction results in terms of exact match accuracy (%).

Methods	Top-k accuracy %			
	1	3	5	10
G2Gs	16.5	27.8	33.1	40.4
GraphRetro	48.3	58.4	60.5	62.4
Transformer	55.8	70.3	74.8	78.9
Retrosim	56.5	65.8	69.0	73.1
Megan	59.5	73.9	77.9	81.7
Neuralsym	**63.0**	73.3	76.0	78.6
GLN	62.9	**74.1**	**78.4**	**82.7**

6. Case Study



CONTEXT-ENRICHED MOLECULE REPRESENTATIONS IMPROVE FEW-SHOT DRUG DISCOVERY

Johannes Schimunek[1], **Philipp Seidl**[1], **Lukas Friedrich**[2], **Daniel Kuhn**[2], **Friedrich Rippmann**[2], **Sepp Hochreiter**[1], **and Günter Klambauer**[1]

[1] ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University Linz, Austria

schimunek@ml.jku.at

[2] Computational Chemistry & Biologics, Merck Healthcare, Darmstadt, Germany

1. Introduction

- Importance of Few-Shot learning.
- Inadequacy of Current methods.
- Novelty – Associative Memories.

- (Contribution) New naïve baseline.
- (Contribution) SOTA performance.
- (Contribution) Add a new method of enriching representations.

2. Problem

- $y = g_{\boldsymbol{w}}(\boldsymbol{m})$ where w represents the parameters of the network to be trained.

3. Model

- Overview

$$
\begin{aligned}
\text{context module:} \quad & \boldsymbol{m}' = f^{\mathrm{CM}}(\boldsymbol{m}, \boldsymbol{C}) \\
& \boldsymbol{X}' = f^{\mathrm{CM}}(\boldsymbol{X}, \boldsymbol{C}), \\
\text{cross-attention module:} \quad & [\boldsymbol{m}'', \boldsymbol{X}''] = f^{\mathrm{CAM}}([\boldsymbol{m}', \boldsymbol{X}']), \\
\text{similarity module:} \quad & \hat{y} = f^{\mathrm{SM}}(\boldsymbol{m}'', \boldsymbol{X}'', \boldsymbol{y}),
\end{aligned}
$$

$$
\underset{\substack{\text{symbolic or} \\ \text{low-level repr.}}}{m} \xrightarrow{f^{\mathrm{ME}}} \underset{\substack{\text{molecule} \\ \text{embedding}}}{\boldsymbol{m}} \xrightarrow{f^{\mathrm{CM}}} \underset{\substack{\text{context} \\ \text{repr.}}}{\boldsymbol{m}'} \xrightarrow{f^{\mathrm{CAM}}} \underset{\substack{\text{similarity} \\ \text{repr.}}}{\boldsymbol{m}''},
$$

$$
\underset{\substack{\text{symbolic or} \\ \text{low-level repr.}}}{x_n} \xrightarrow{f^{\mathrm{ME}}} \underset{\substack{\text{molecule} \\ \text{embedding}}}{\boldsymbol{x}_n} \xrightarrow{f^{\mathrm{CM}}} \underset{\substack{\text{context} \\ \text{repr.}}}{\boldsymbol{x}'_n} \xrightarrow{f^{\mathrm{CAM}}} \underset{\substack{\text{similarity} \\ \text{repr.}}}{\boldsymbol{x}''_n}.
$$

- Context Module (CM)
  - *Hopfield Layer*, ($\boldsymbol{\Xi}$ and $\boldsymbol{C}$ can have different dimensions)

$$
\mathrm{Hopfield}(\boldsymbol{\Xi}, \boldsymbol{C}) := (\boldsymbol{W}_E \boldsymbol{C})\, \mathrm{softmax}\left(\beta\, (\boldsymbol{W}_C \boldsymbol{C})^T\, (\boldsymbol{W}_\Xi \boldsymbol{\Xi})\right)
$$

  - Forward pass,

$$
\boldsymbol{m}' = \mathrm{Hopfield}(\boldsymbol{m}, \boldsymbol{C}) \quad \text{and} \quad \boldsymbol{X}' = \mathrm{Hopfield}(\boldsymbol{X}, \boldsymbol{C})
$$

3. Model

- Cross Attention Module (CAM)
  - Forward pass, (this is exactly a normal attention module)

$$[\boldsymbol{m}'', \boldsymbol{X}''] = \text{Hopfield}([\boldsymbol{m}', \boldsymbol{X}'], [\boldsymbol{m}', \boldsymbol{X}'])$$

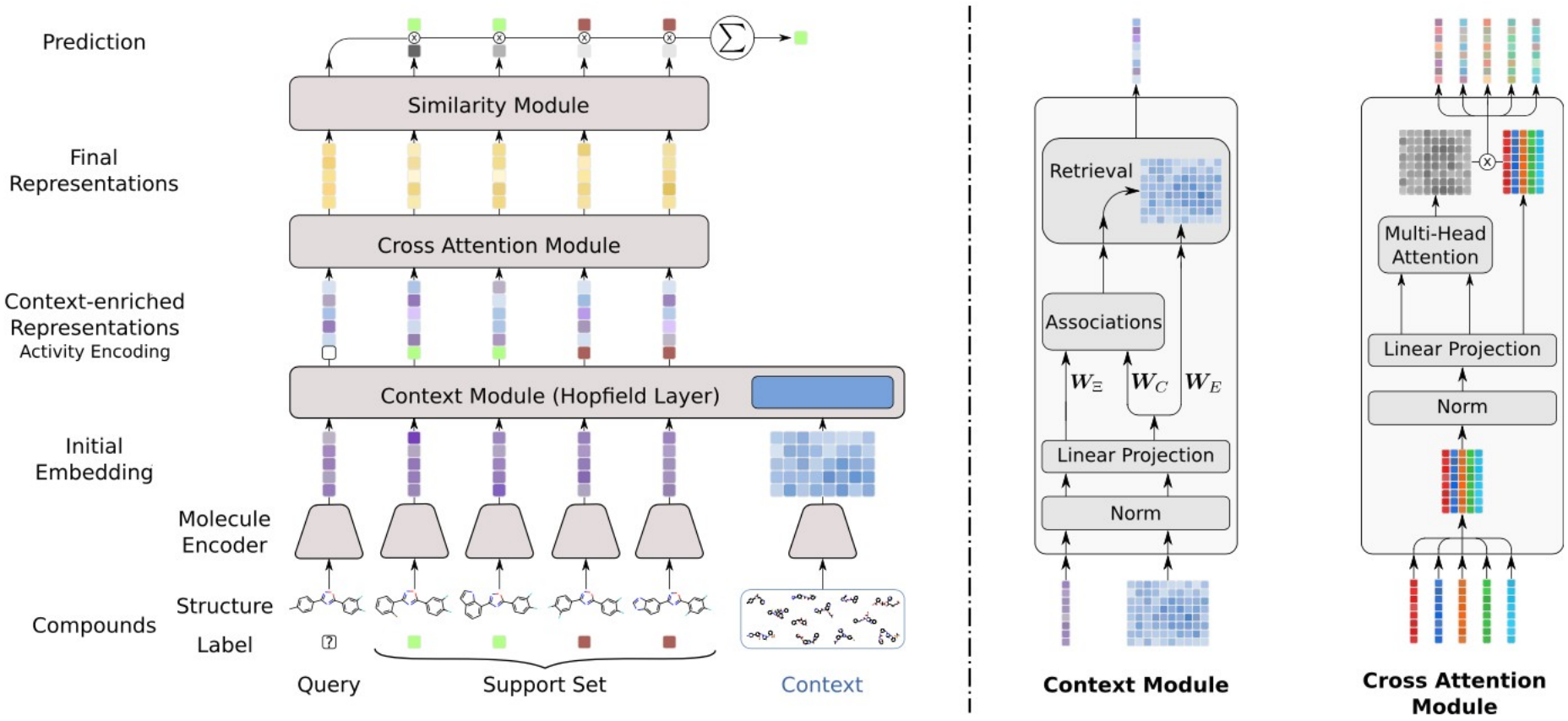
- Similarity Module (SM)

$$\hat{y} = \sigma\left(\tau^{-1}\frac{1}{N}\sum_{n=1}^{N} y'_n\, k(\boldsymbol{m}'', \boldsymbol{x}''_n)\right)$$

- Balancing strategy for labels

$$y'_n = \begin{cases} N/(\sqrt{2N_A}) & \text{if } y_n = 1 \\ -N/(\sqrt{2N_I}) & \text{else} \end{cases}$$

3. Model


Prediction
Similarity Module
Final Representations
Cross Attention Module
Context-enriched Representations
Activity Encoding
Context Module (Hopfield Layer)
Initial Embedding
Molecule Encoder
Compounds
Structure
Label
Query
Support Set
Context
Retrieval
Associations
W_Ξ
W_C
W_E
Linear Projection
Norm
Context Module
Multi-Head Attention
Linear Projection
Norm
Cross Attention Module

4. Related Work

- Nothing special but they have cited many papers from 2014-2018, but not many recent ones, weird.

5. Experiments

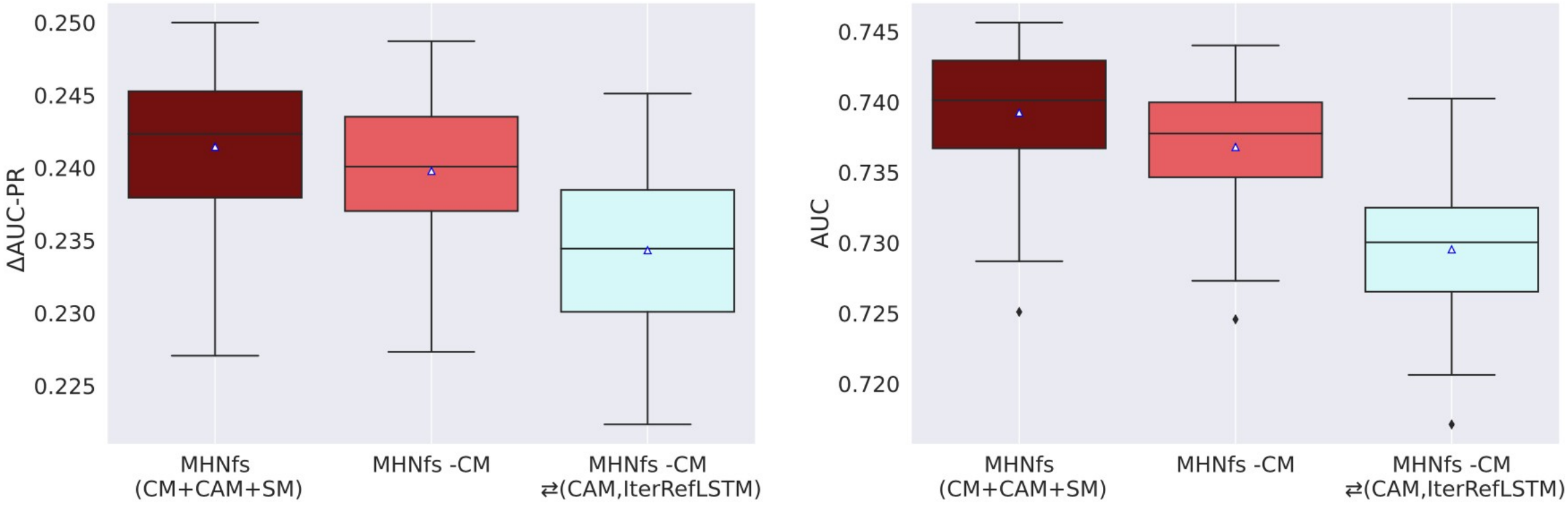
Method	All [157]	Kin. [125]	Hydrol. [20]	Oxid.[7]
GNN-ST[a] (Stanley et al., 2021)	.029 ± .004	.027 ± .004	.040 ± .018	.020 ± .016
MAT[a] (Maziarka et al., 2020)	.052 ± .005	.043 ± .005	.095 ± .019	.062 ± .024
Random Forest[a] (Breiman, 2001)	.092 ± .007	.081 ± .009	.158 ± .028	.080 ± .029
GNN-MT[a] (Stanley et al., 2021)	.093 ± .006	.093 ± .006	.108 ± .025	.053 ± .018
Similarity Search	.118 ± .008	.109 ± .008	.166 ± .029	.097 ± .033
GNN-MAML[a] (Stanley et al., 2021)	.159 ± .009	.177 ± .009	.105 ± .024	.054 ± .028
PAR (Wang et al., 2021)	.164 ± .008	.182 ± .009	.109 ± .020	.039 ± .008
Frequent Hitters	.182 ± .010	.207 ± .009	.098 ± .009	.041 ± .005
ProtoNet[a] (Snell et al., 2017)	.207 ± .008	.215 ± .009	.209 ± .030	.095 ± .029
Siamese Networks (Koch et al., 2015)	.223 ± .010	.241 ± .010	.178 ± .026	.082 ± .025
IterRefLSTM (Altae-Tran et al., 2017)	.234 ± .010	.251 ± .010	.199 ± .026	.098 ± .027
ADKF-IFT[b] (Chen et al., 2022)	.234 ± .009	.248 ± .020	**.217** ± .017	**.106** ± .008
MHNfs (ours)	**.241** ± .009	**.259** ± .010	.199 ± .027	.096 ± .019

5. Experiments

- 2 new baselines → *Frequent Hitters* and *Similarity Search.*

$$\hat{y} = 1/N \sum_{n=1}^{N} y_n \; k(\boldsymbol{m}, \boldsymbol{x}_n)$$

- Ablations



5. Experiments

- Domain Shift test

Method	AUC	ΔAUC-PR
Similarity Search (baseline)	.629 ± .015	.061 ± .008
IterRefLSTM (Altae-Tran et al., 2017)	.664 ± .018	.067 ± .008
MHNfs (ours)	**.679** ± .018	**.073** ± .008

6. Inference

- This feels like making the notion of a knowledge graph very concrete, can be thought as if during inference/training the method has access to the full knowledge graph.