

A Fast and Interpretable Context-aware Parser for Isolated Formulas

Ayush Kumar Shah

Advisor : **Dr. Richard Zanibbi**

{as1211, rxzvcs}@rit.edu

R·I·T

B. THOMAS GOLISANO
College of **COMPUTING AND**
INFORMATION SCIENCES



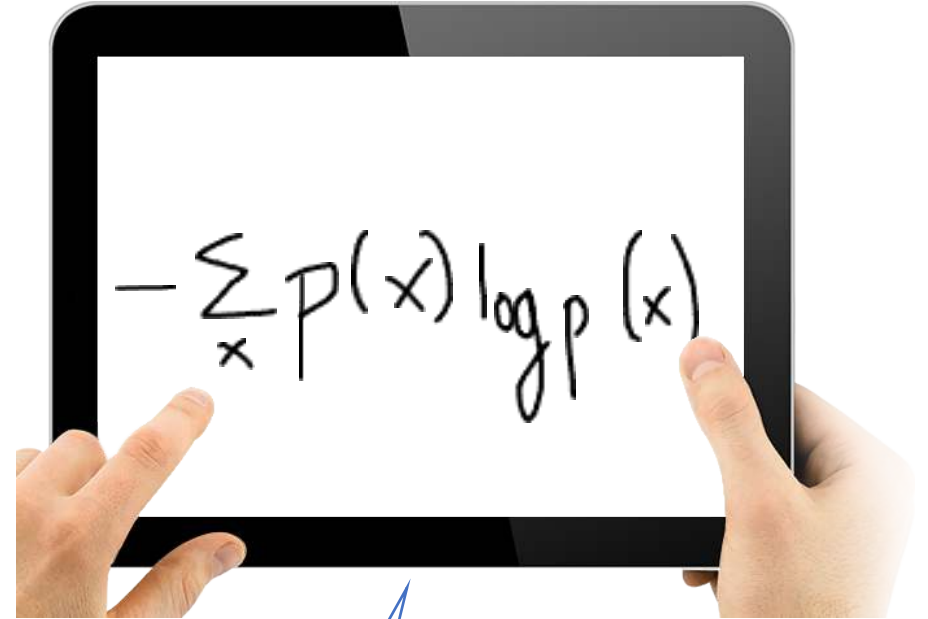
Document and Pattern Recognition Lab

Motivation: Information Retrieval

$$-\sum_x p(x) \log p(x)$$



What is the name
of this formula?



Let's find documents
describing this formula.

Motivation: Inf. Retrieval - Question Answering

What is the value of
this integral?

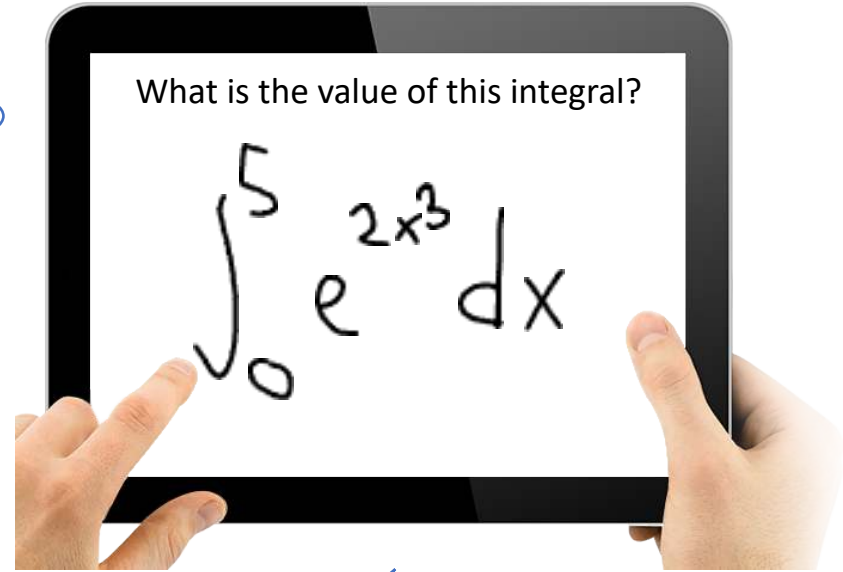
$$\int_0^5 e^{2x^3} dx$$

How is this
computed?

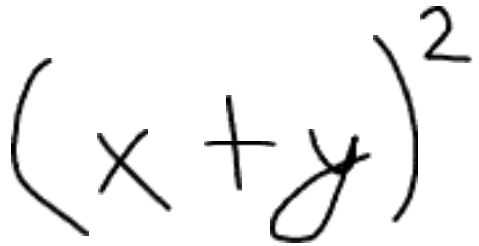
What is the value of this integral?

$$\int_0^5 e^{2x^3} dx$$

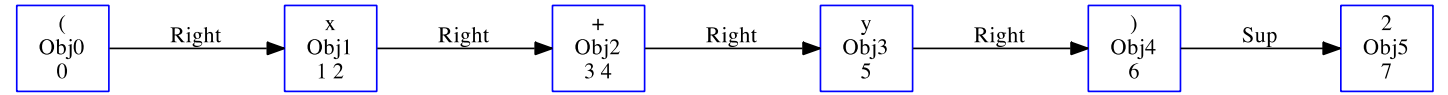
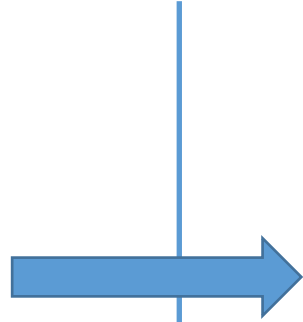
How big is the value of
this integral?



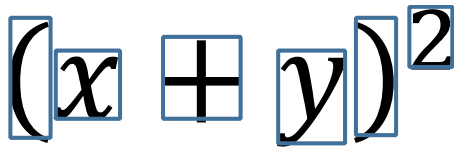
Inputs and Graphs Describing Appearance



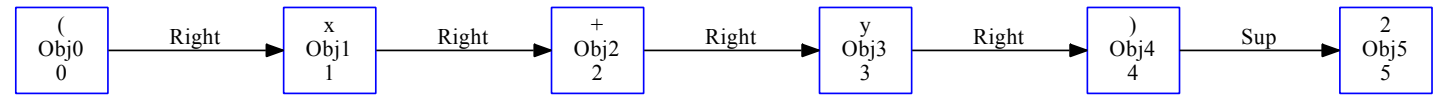
Offline / Online handwritten strokes



Symbol Layout Tree (SLT) over Symbols

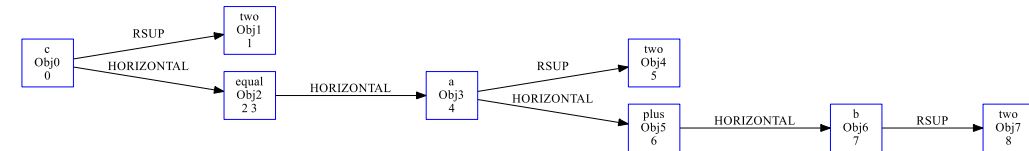
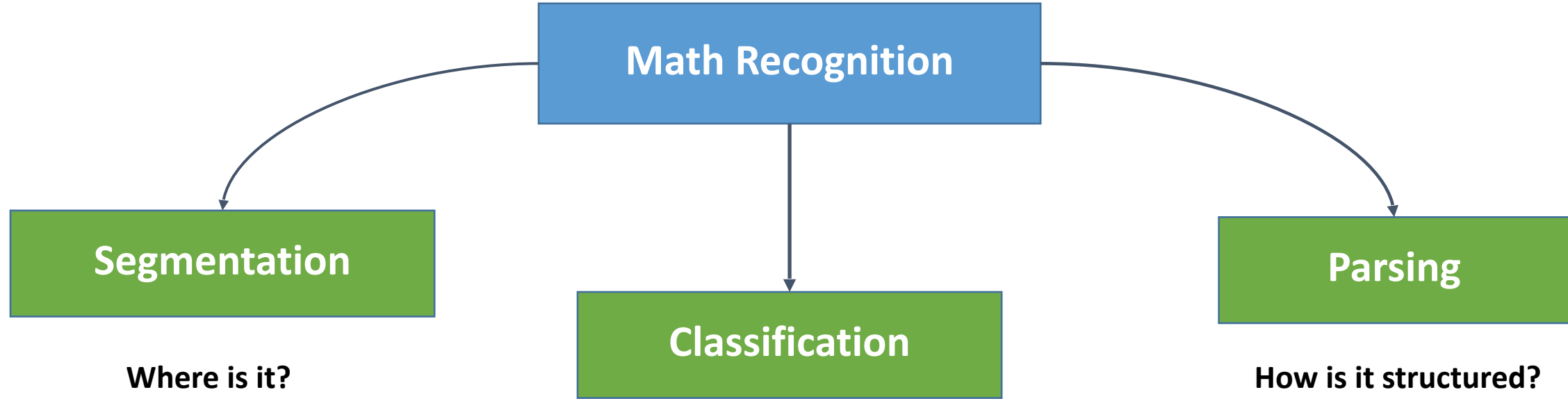


Typeset formula image



Symbol Layout Tree over Connected Components (= Symbols)

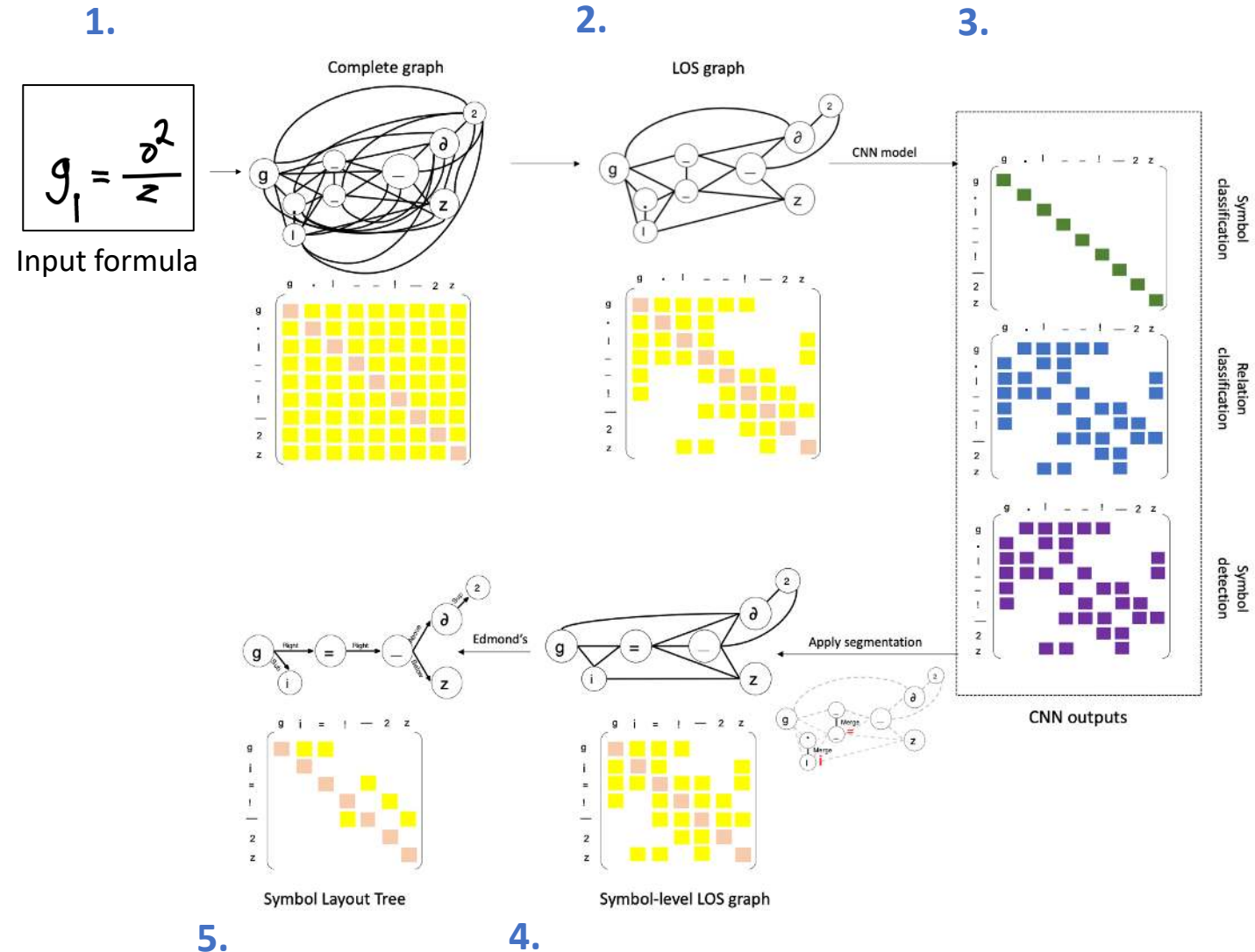
The Recognition Problem



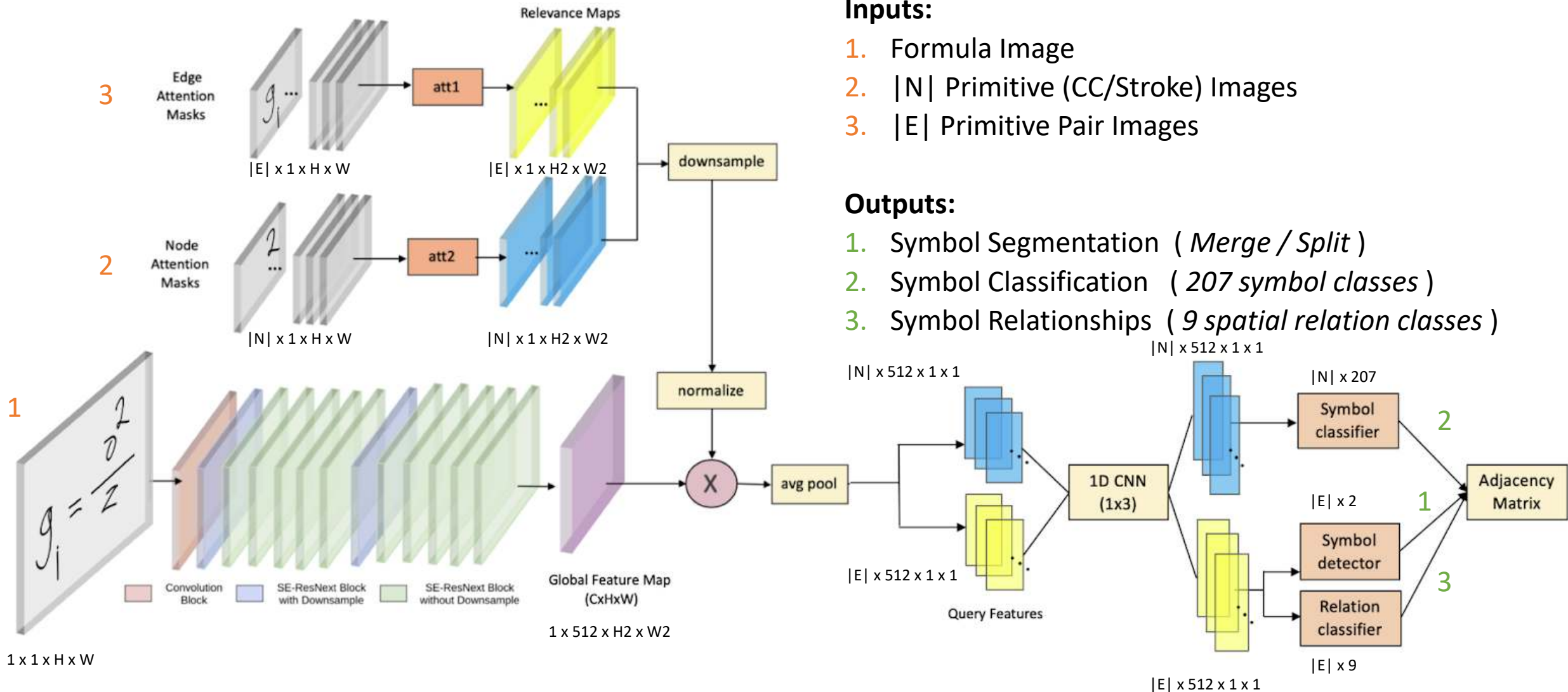
Symbol + Relationship labels

QD-GGA: Query-Driven Global Graph Attention Parser

1. **Construct** graph over CCs/strokes
2. **Prune**: Convert to LOS graph
3. **Classify** edges as merge/split and relationships, nodes as symbols
4. **New LOS graph**: detected symbols
5. **Extract MST** using Edmond's arborescence algorithm



QD-GGA: Query-Driven Global Graph Attention Parser



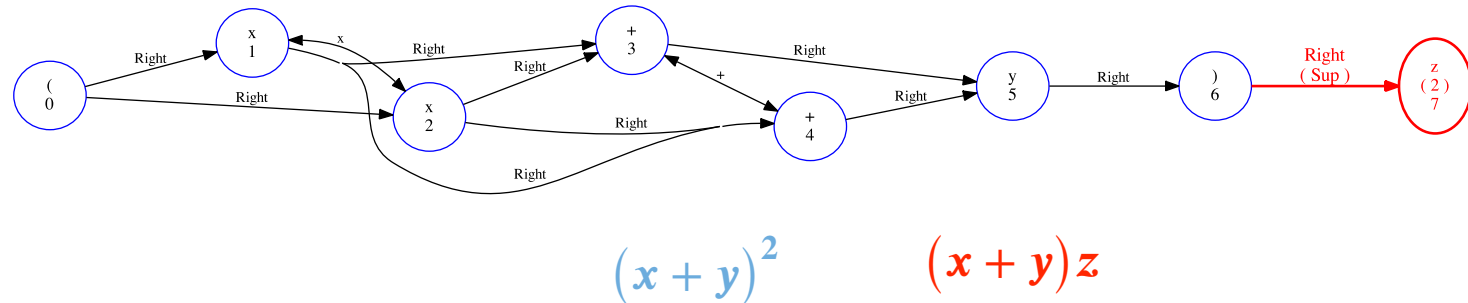
Experiments

1. Training and testing acceleration through

- Distributed Data Parallel (DDP) using multiple GPUs
- Custom data loader with dynamic batch size (as per size of formula inputs on GPU)
- More than 5 times speedup in training and 6 times speedup in testing

Evaluation results on InftyMCCDB-2¹ test set (6830 formula images) using models trained on 12,551 formula images

Experiment Configuration (Model)	Context module (conv1d)	# of GP Us	Batch Size per GPU	Epochs	Symbols		Relationships		Formulas		Training Time per epoch
					Detection	Det. + Class	Detection	Det. + Class	Structure	Str. + Class	
Infty_contour	Yes	1	1	89	96.68	94.17	95.06	94.71	80.26	73.32	1 hour 45 mins
Infty_contour	Yes	4	Dynamic	89	96.46	92.93	93.99	93.55	79.17	70.16	20 mins



¹ <https://www.cs.rit.edu/~dprl/data/InftyMCCDB-2/>

QD-GGA: Error Analysis

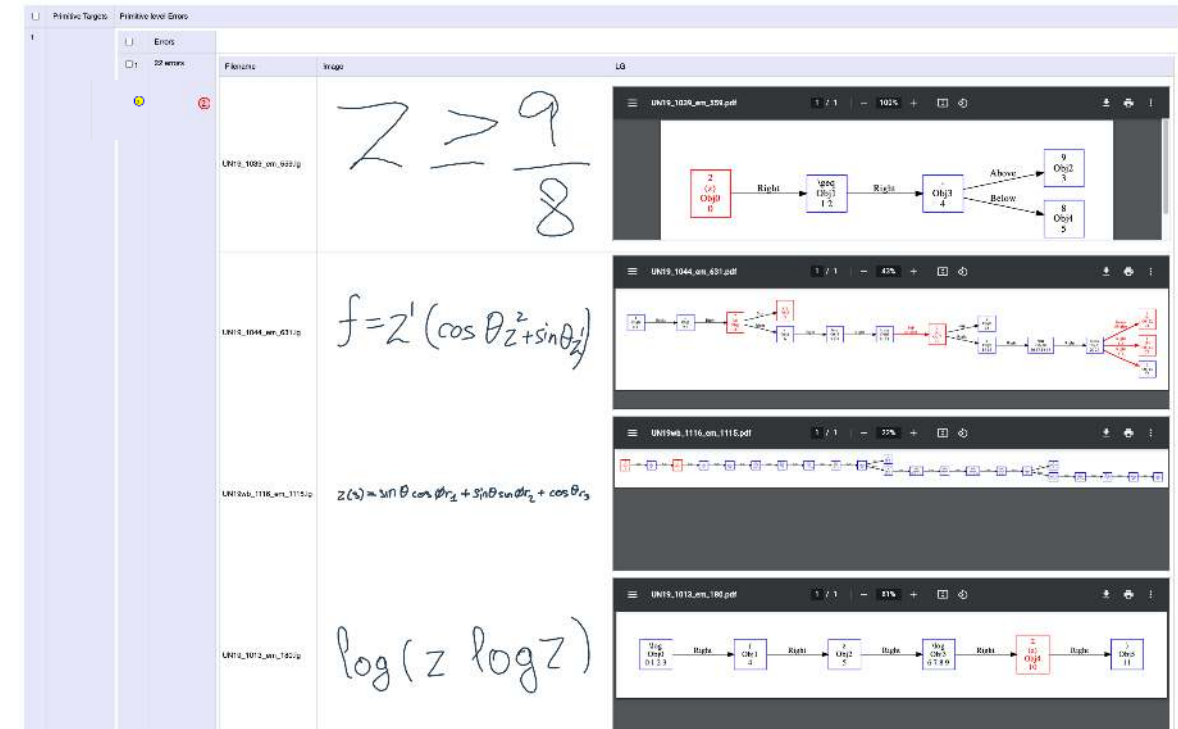
ConfHist Outputs for CROHME (strokes data)

Object Confusion Histograms

Object structures recognized incorrectly are shown at left, sorted by decreasing frequency. 95 incorrect targets, 1418 errors.

Object Targets		Primitive Targets and Errors							
3	55 errors z	☐ Targets							
		☐ 1 51 errors z	☐ 22 errors 2	☐ 7 errors y	☐ 5 errors g	☐ 4 errors x	☐ 3 errors 1		
		☐ 2 4 errors z z	☐ 2 errors 2 2	☐ 2 errors 1 1					

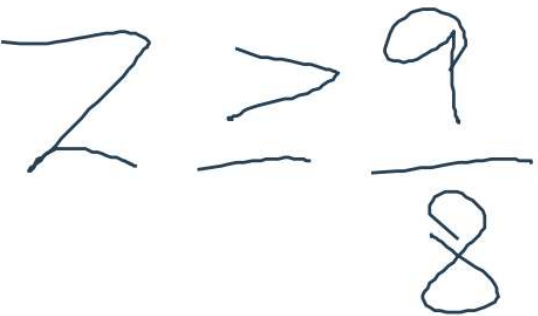
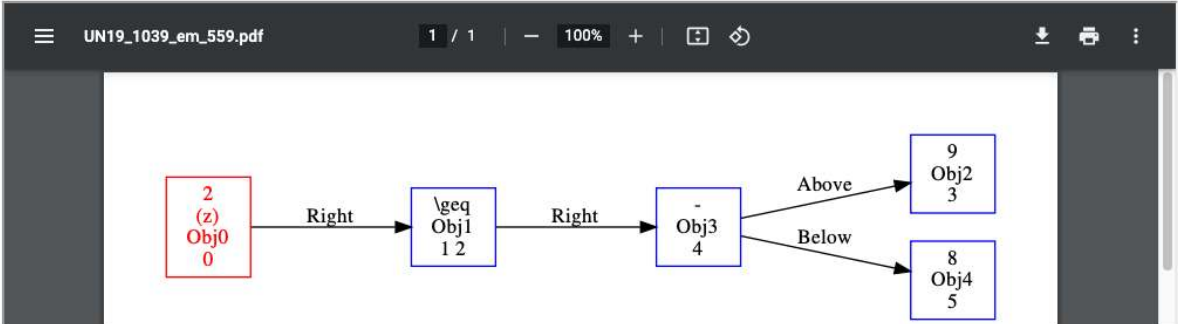
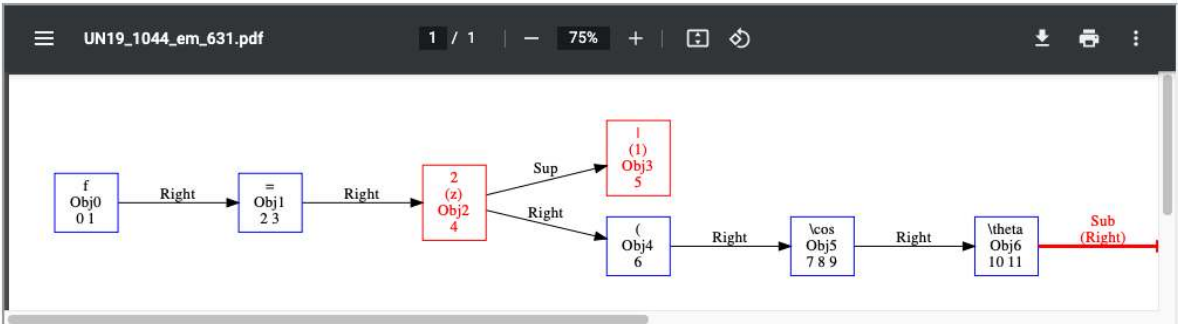
Errors organized by decreasing frequency



Specific instances where 'z' is misclassified as '2',
seen after clicking on the '22 errors' link

QD-GGA: Error Analysis

ConfHist Outputs for CROHME (strokes data)

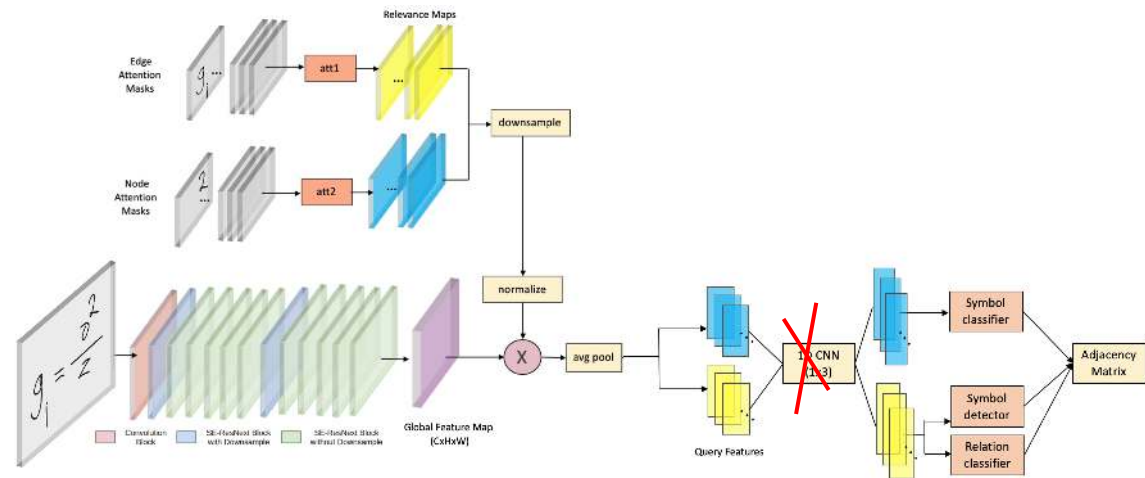
☐ Primitive Targets		Primitive level Errors		
1	z	☐ Errors		
		☐ 1 22 errors		
		Filename	Image	LG
		UN19_1039_em_559.lg		
		UN19_1044_em_631.lg		

Zoomed in: Specific instances where 'z' is misclassified as '2,' seen after clicking on the '22 errors' link

Experiments

2. Removed the context module (1D convolution layers at the end)

- The detection and classification rates decreased by a huge factor
- Shows importance of context



Evaluation results on InftyMCCDB-2¹ test set (6830 formula images) using models trained on 12,551 formula images

Experiment Configuration (Model)	Context module (conv1d)	# of GP Us	Batch Size per GPU	Epochs	Symbols		Relationships		Formulas		Training Time per epoch
					Detection	Det. + Class	Detection	Det. + Class	Structure	Str. + Class	
Infty_contour	Yes	1	1	89	96.68	94.17	95.06	94.71	80.26	73.32	1 hour 45 mins
Infty_contour	Yes	4	Dynamic	89	96.46	92.93	93.99	93.55	79.17	70.16	20 mins
Infty_no_conv1d	No	4	Dynamic	89	91	6.76	67.62	53.57	48.16	13.84	20 mins

¹ <https://www.cs.rit.edu/~dprl/data/InftyMCCDB-2/>

Experiments

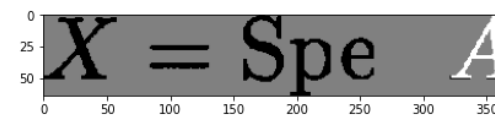
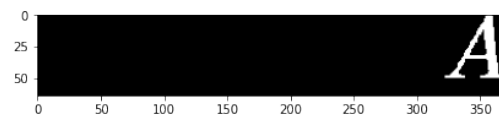
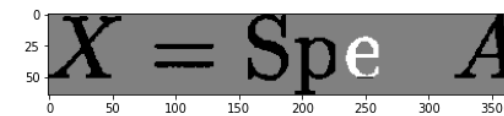
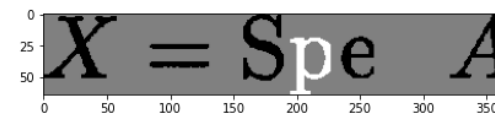
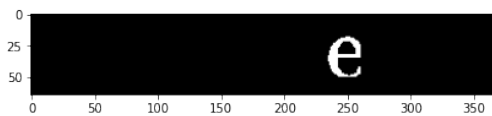
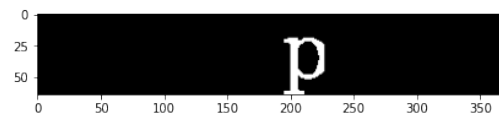
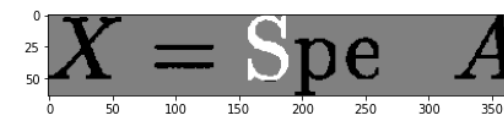
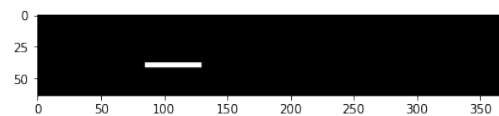
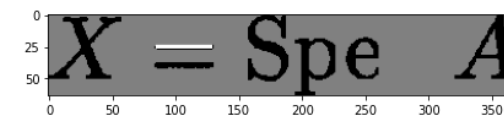
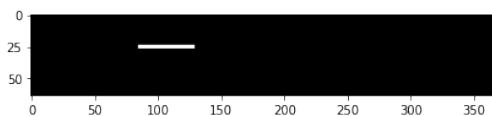
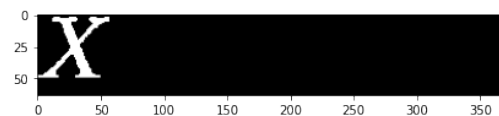
3. Modify the feature representations to increase global context

$X = \text{Spe } A$

Input (28000016.png)

■ -1 for context

□ +1 for query



Old CCs

Modified CCs

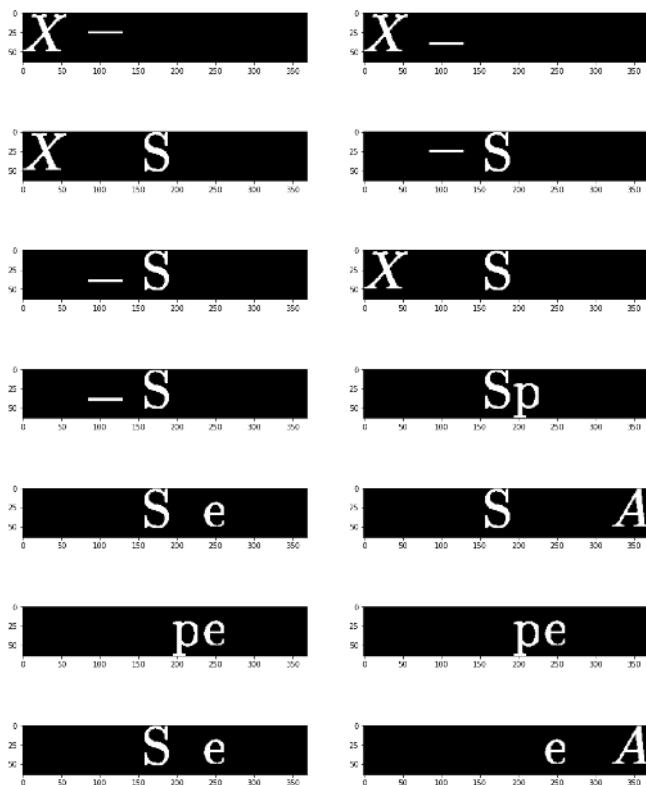
Experiments

3. Modify the feature representations to increase global context

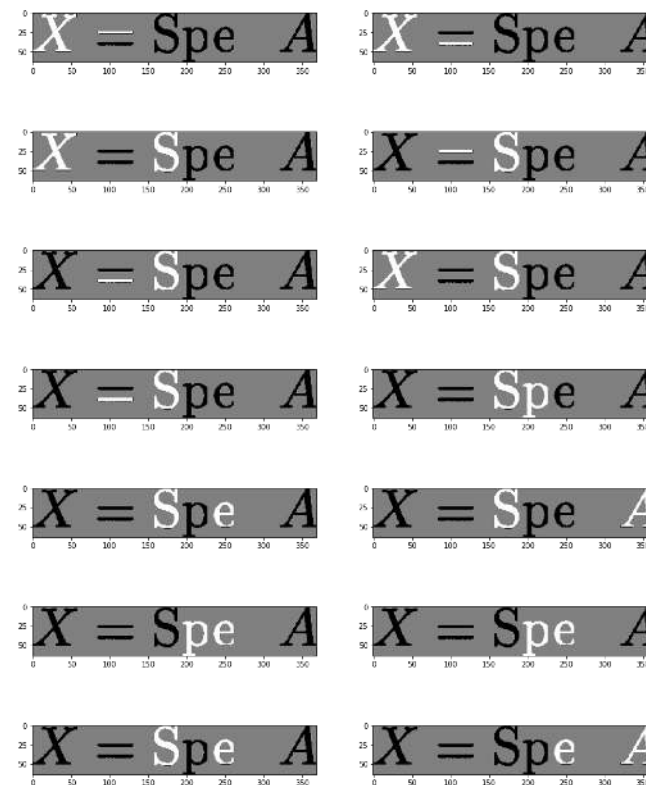
$X = \text{Spe } A$

Input (28000016.png)

■ -1 for context
□ +1 for query



Old CC pairs (LOS edges)



Modified CC pairs (LOS edges)

Experiments

3. Modify the feature representations (+1, -1) to increase global context

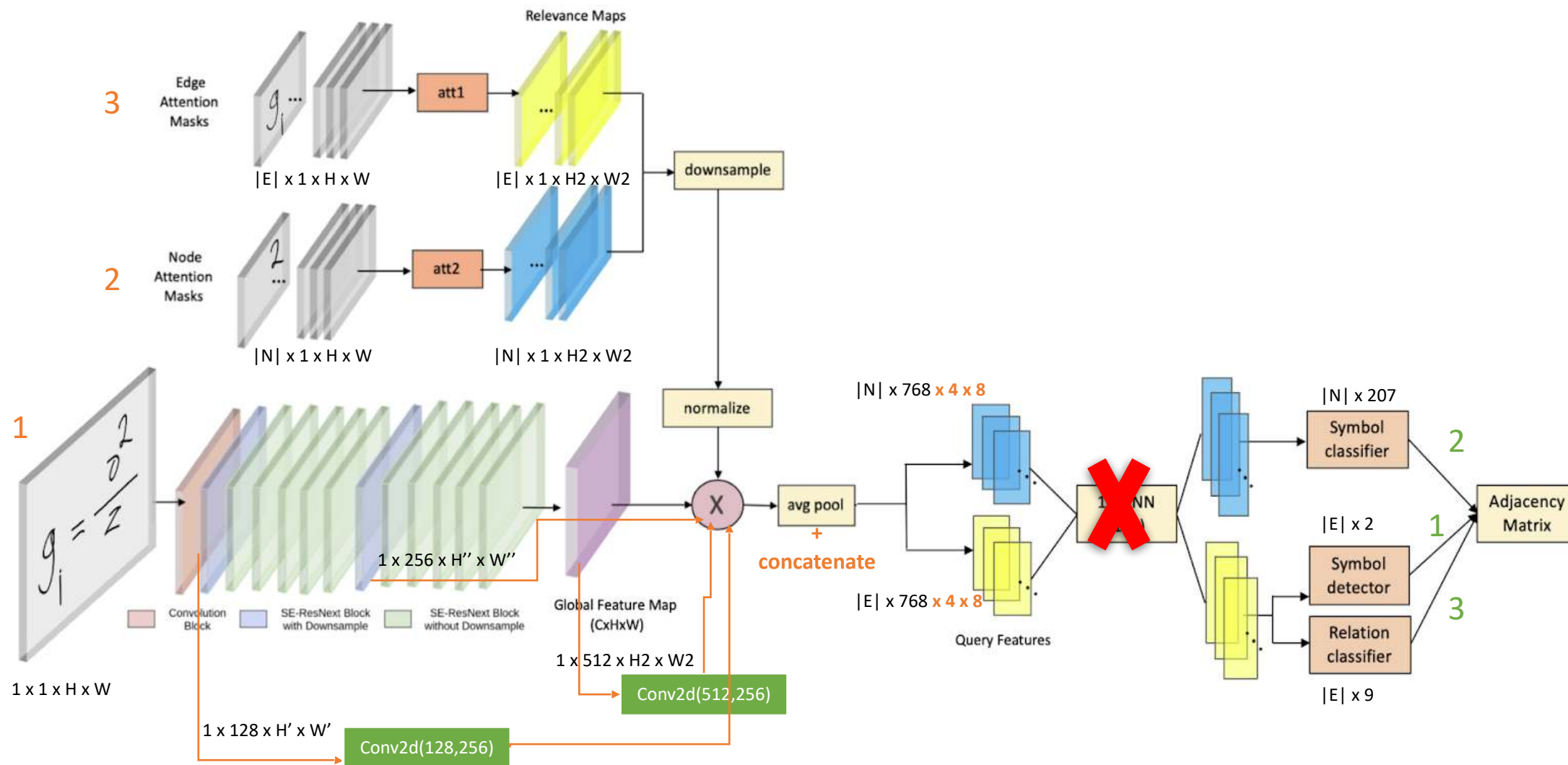
- The scores decreased. Possible causes might be:
 - more variance of context in the features,
 - average pool punishes the approach since (-1) would decrease the average

Evaluation results on InftyMCCDB-2¹ test set (6830 formula images) using models trained on 12,551 formula images

Experiment Configuration	Context module (conv1d)	# of GP Us	Batch Size per GPU	Epochs	Symbols		Relationships		Formulas		Training Time per epoch
					Detection	Det. + Class	Detection	Det. + Class	Structure	Str. + Class	
Infty_contour	Yes	1	1	89	96.68	94.17	95.06	94.71	80.26	73.32	1 hour 45 mins
Infty_contour	Yes	4	Dynamic	89	96.46	92.93	93.99	93.55	79.17	70.16	20 mins
Infty_no_conv1d	No	4	Dynamic	89	91	6.76	67.62	53.57	48.16	13.84	20 mins
Infty_modified_features	Yes	4	Dynamic	89	94.77	75.08	85.92	83.91	69.93	36.15	20 mins

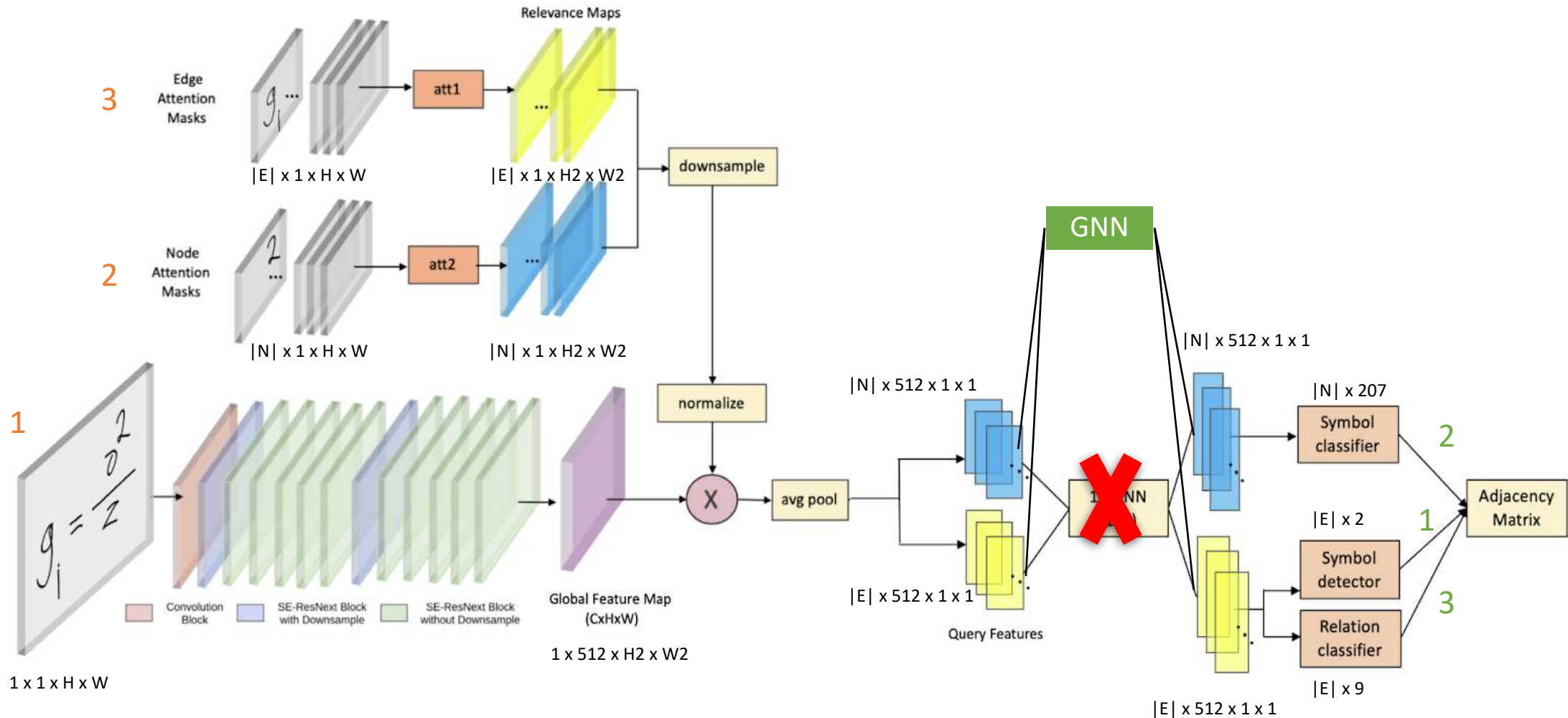
Experiments

4. Add global image features at different resolutions into the attention masks and modify pooling approach to preserve more information **[IN PROGRESS]**



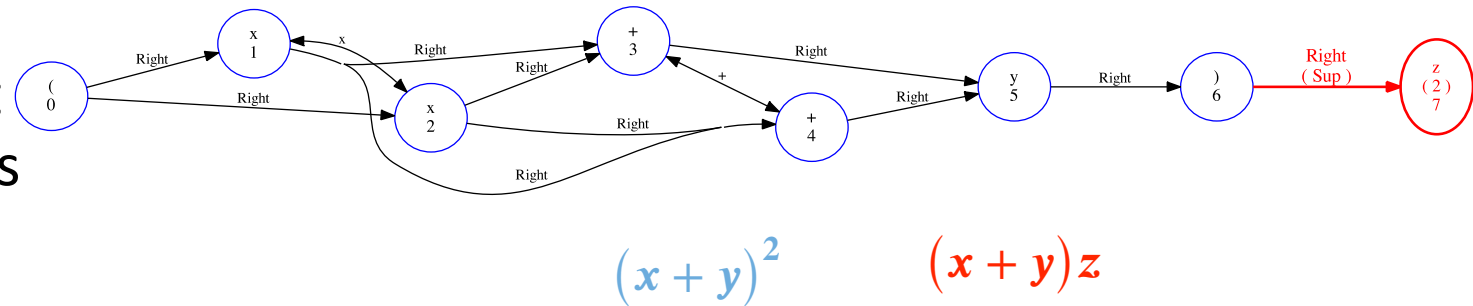
Experiments

- Replace the 1D 'context' convolutions before the output by a Graph Neural Net (GNN) to add better contextual information **[IN PROGRESS]**



Conclusion and Future Work

- Open-source parser for typeset and handwritten math formulas
- High interpretability and fast
- Simple/automatic error analysis
- Data driven i.e.; expression grammars not required



Future work

- Improve global context and use of spatial information in the model
- Match recognition rates with current SOTA models (encoder-decoder)
- Modify QD-GGA for parsing chemical diagrams

Thank You

This material is based on upon work supported by the Alfred P. Sloan Foundation under Grant No. G-2017-9827 and the National Science Foundation (USA) under Grant Nos. IIS-1717997 (MathSeer project) and 2019897 (MMLI project)

Thanks to M. Condon, L. Hu, M. Mahdavi, A. Dey, and M. Langsenkamp for their contributions to the formula extraction and parser



Alfred P. Sloan
FOUNDATION



Q&A