

Line-of-sight with Graph Attention Parser (LGAP) for Math Formulas Recognition

Ayush Kumar Shah, 3rd year PhD

Advisor : **Dr. Richard Zanibbi**

{as1211, rxzvcs}@rit.edu

R·I·T

B. THOMAS GOLISANO
College of COMPUTING AND
INFORMATION SCIENCES



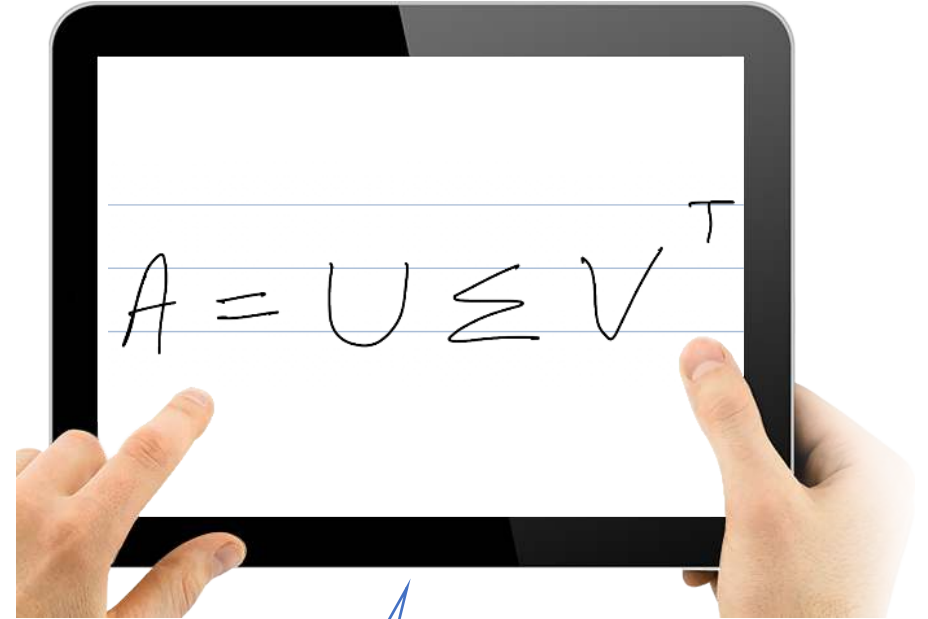
Document and Pattern Recognition Lab

Motivation: Information Retrieval

$$A = U\Sigma V^T$$

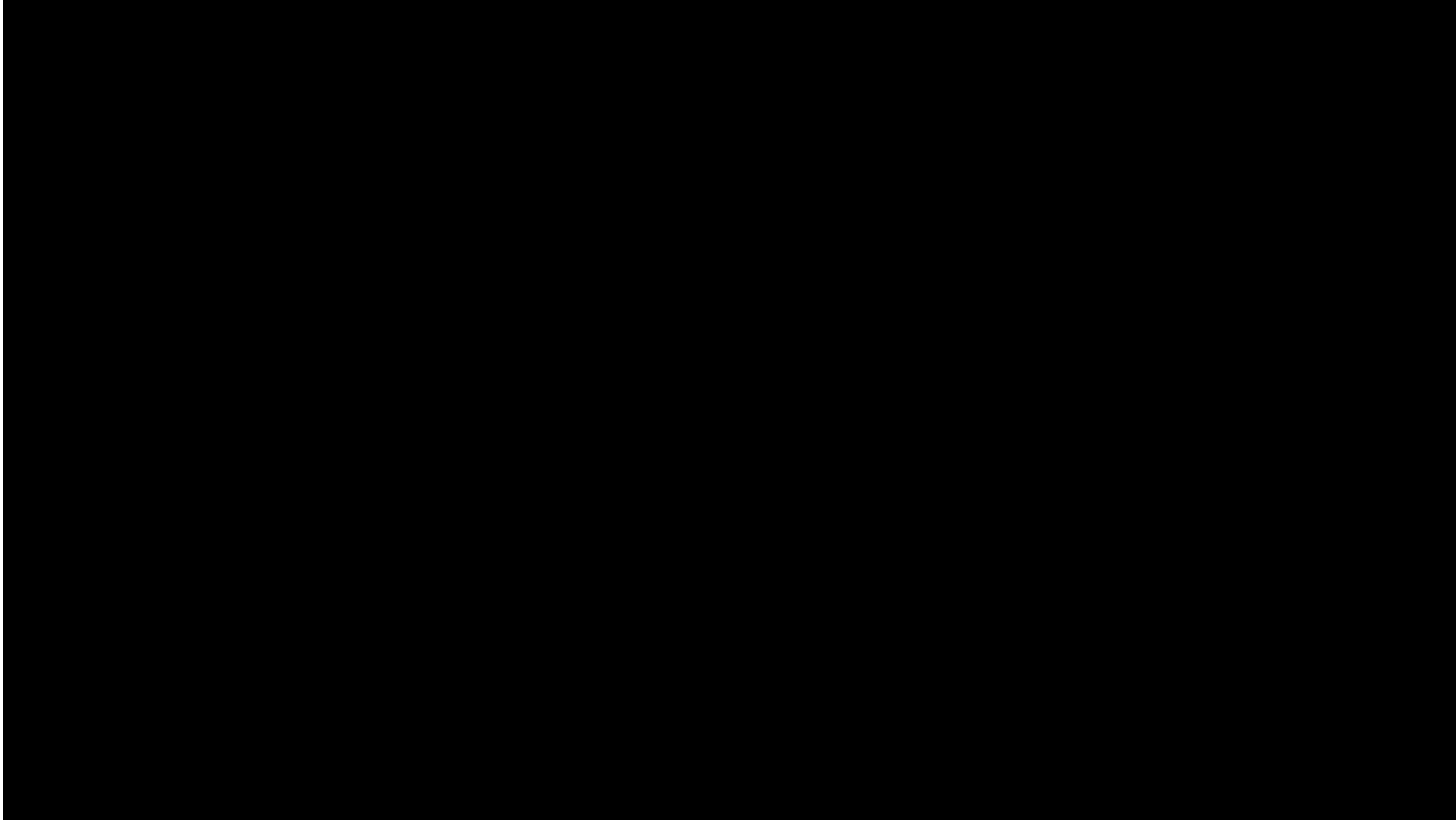


What is the name
of this formula?

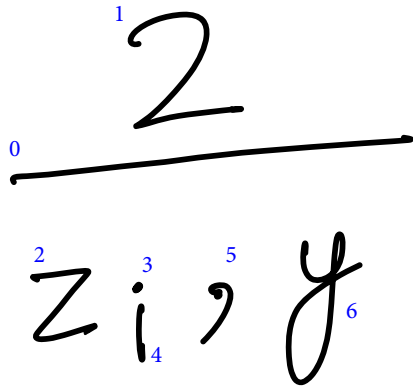


Let's find documents
describing this formula.

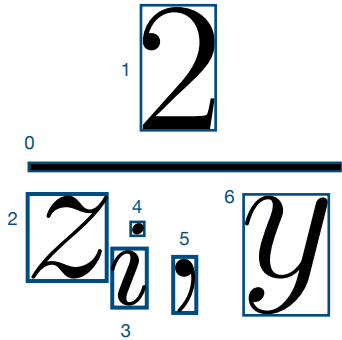
Motivation: Information Retrieval



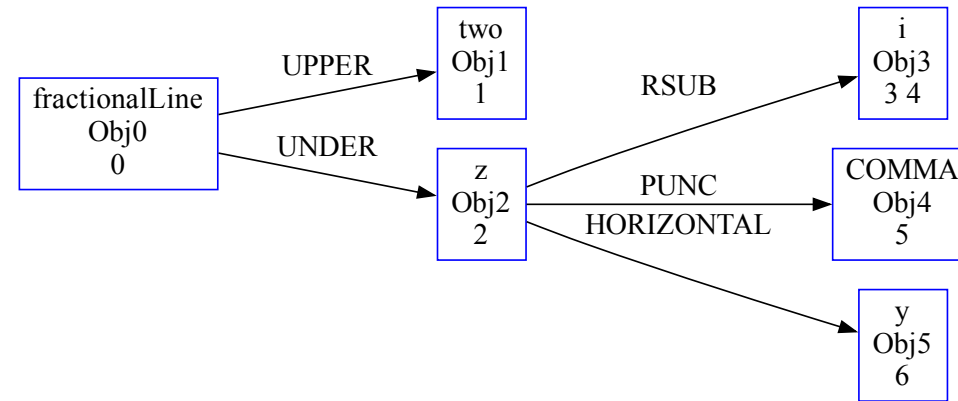
Inputs and Output Graphs Describing Appearance



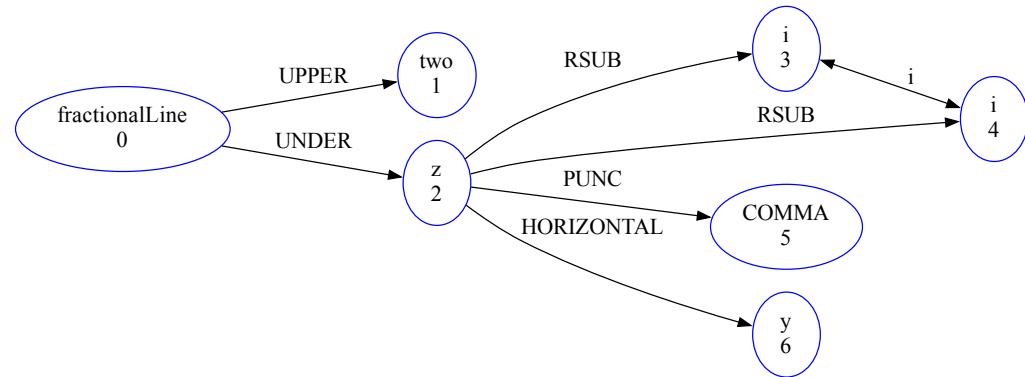
Handwritten strokes



Typeset formula image
(rendered LaTeX)

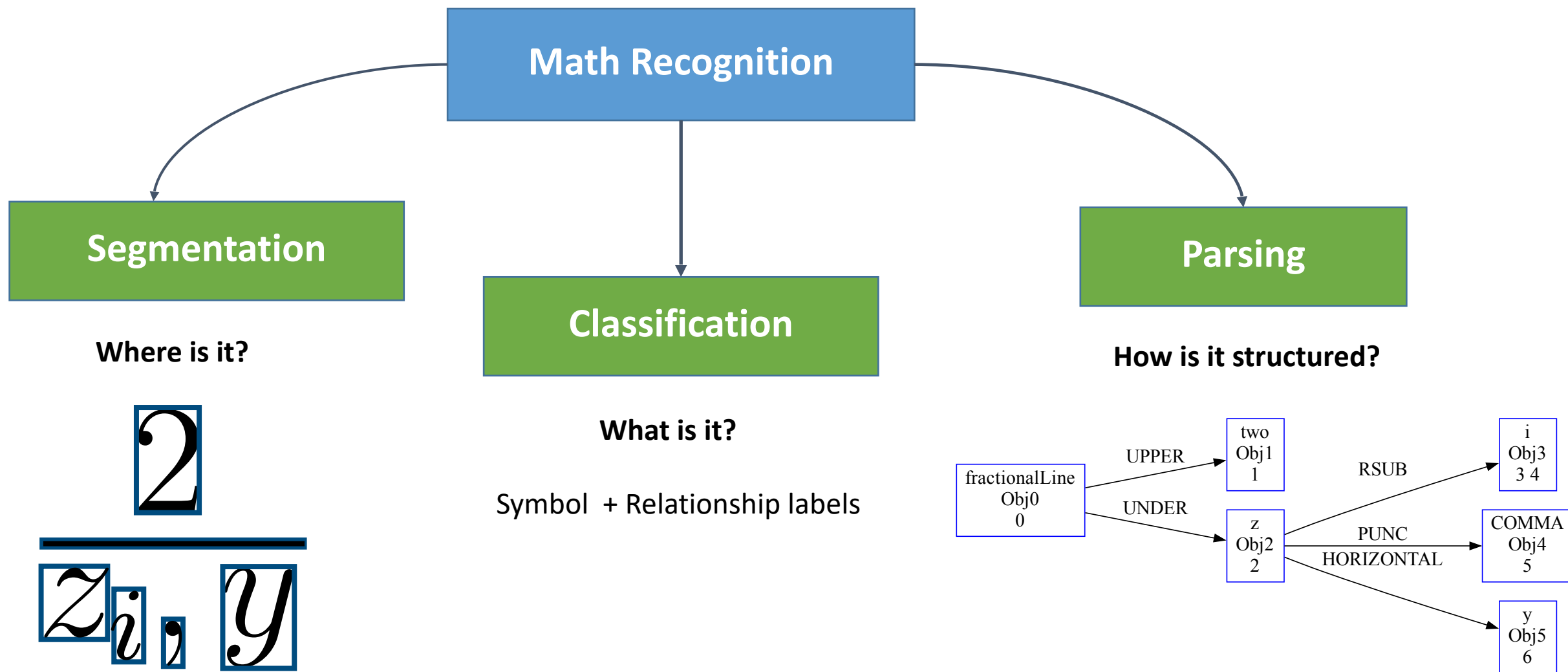


Symbols: Symbol Layout Tree (SLT)



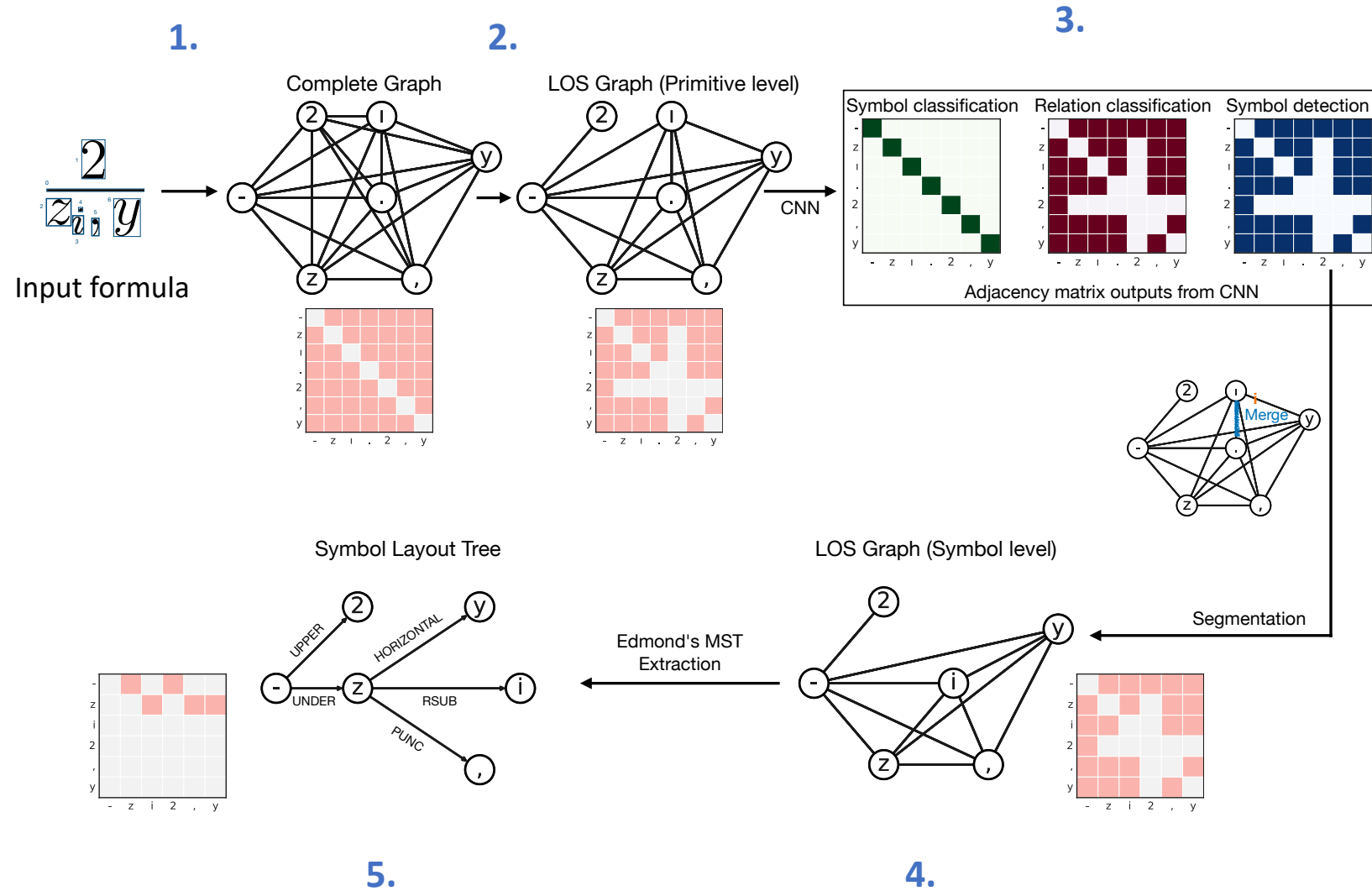
Strokes / Connected Components: Labeled Graph

The Recognition Problem



LGAP Formula Parsing Algorithm

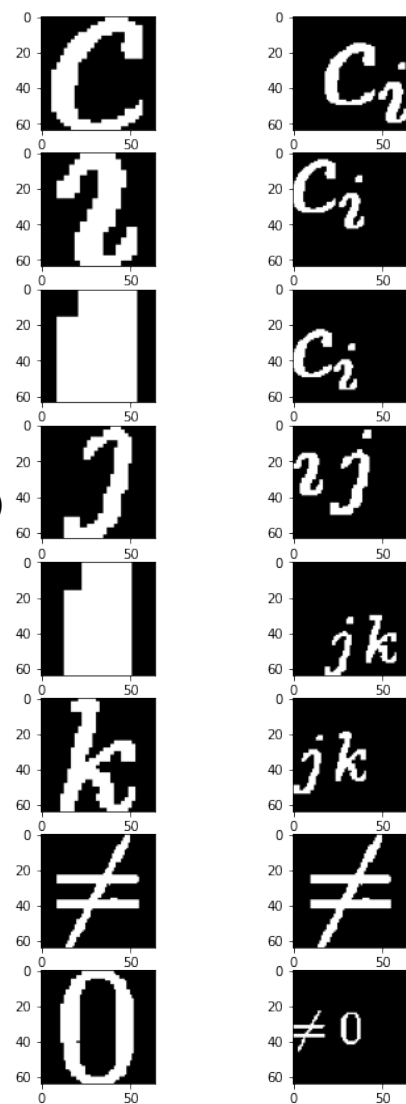
1. **Construct** graph over CCs/ strokes
2. **Prune:** Convert to LOS graph
3. **Classify** edges as merge/split and relationships, nodes as symbols
4. **New LOS graph:** detected symbols
5. **Extract MST** using Edmond's arborescence algorithm



Improvements: Visual Features (Typeset)

$$c_{ijk} \neq 0$$

Shape Features Context Features



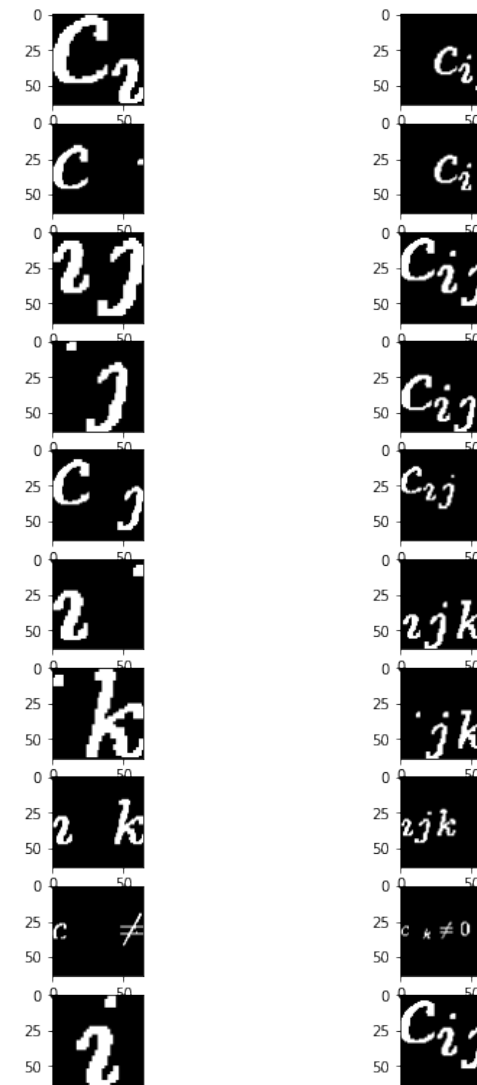
<Cropped CC>

<CC+2NN>
(Centered on parent)

Edge Features
(Segmentation and
Relation Classification)

Shape Features

Context Features

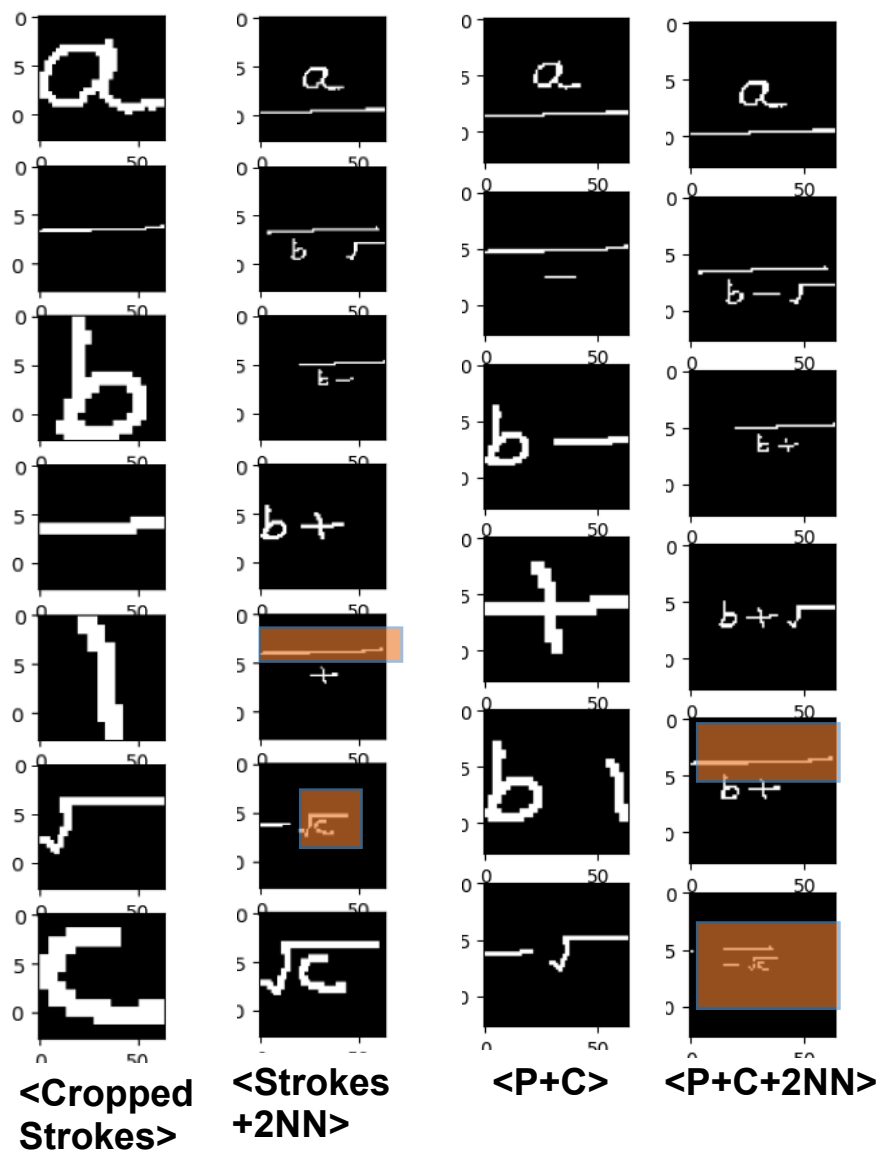
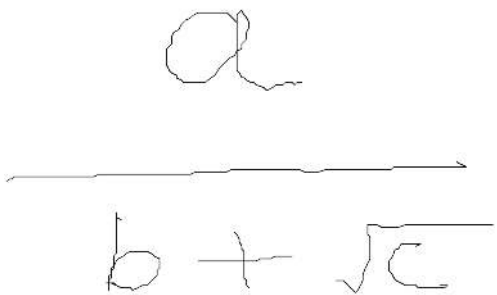


<Parent + Child
CCs (Cropped)>

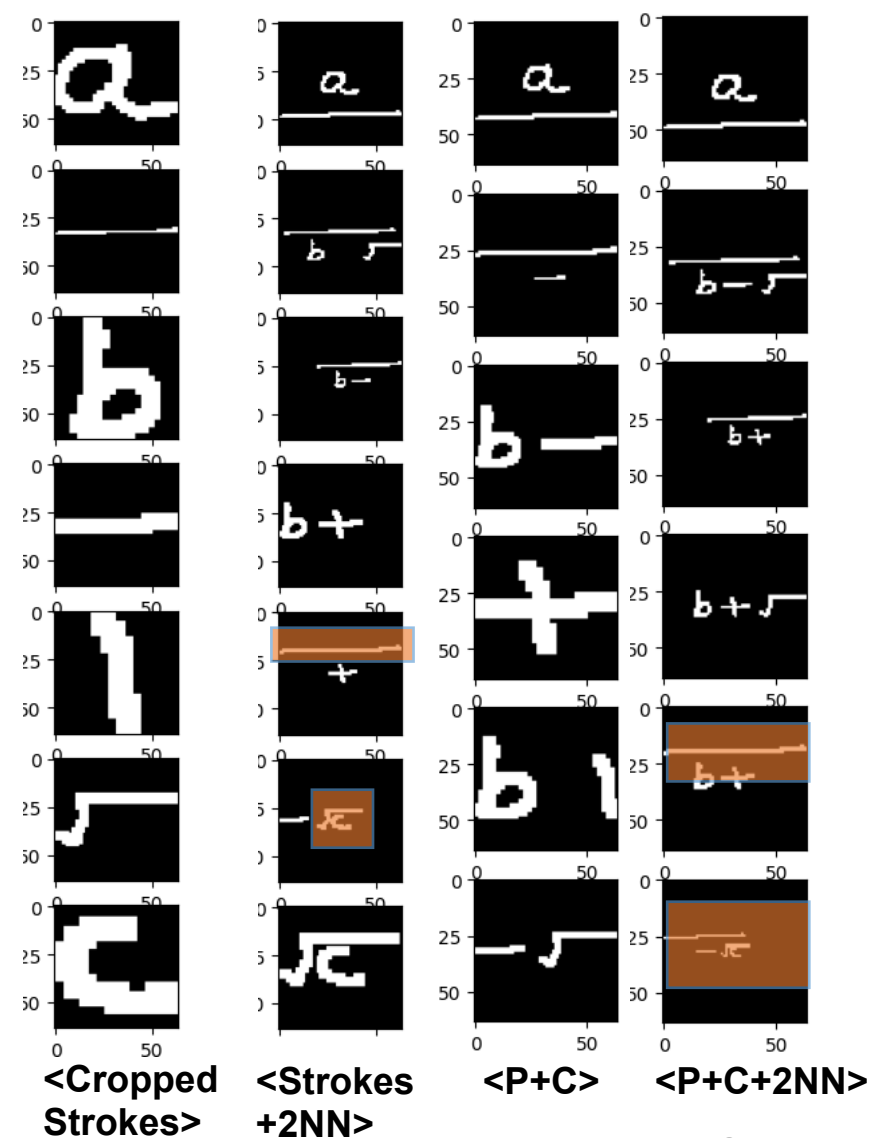
<Parent + Child CCs + 2NN>
(Centered on parent)

Improvements: Visual Features (Handwritten)

Dilation effect for
handwritten dataset:



Dilation: iterations = 1



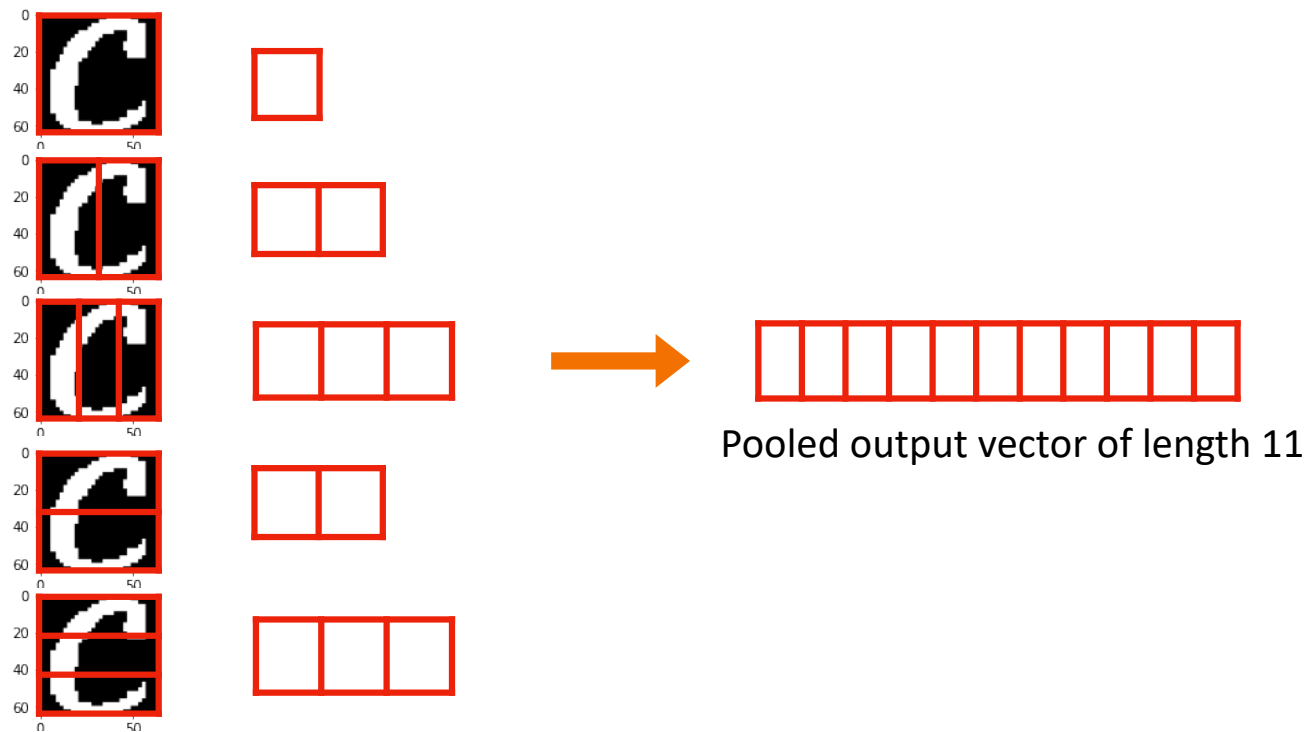
Dilation: iterations = 2

Improvements: Visual Features (SPP)

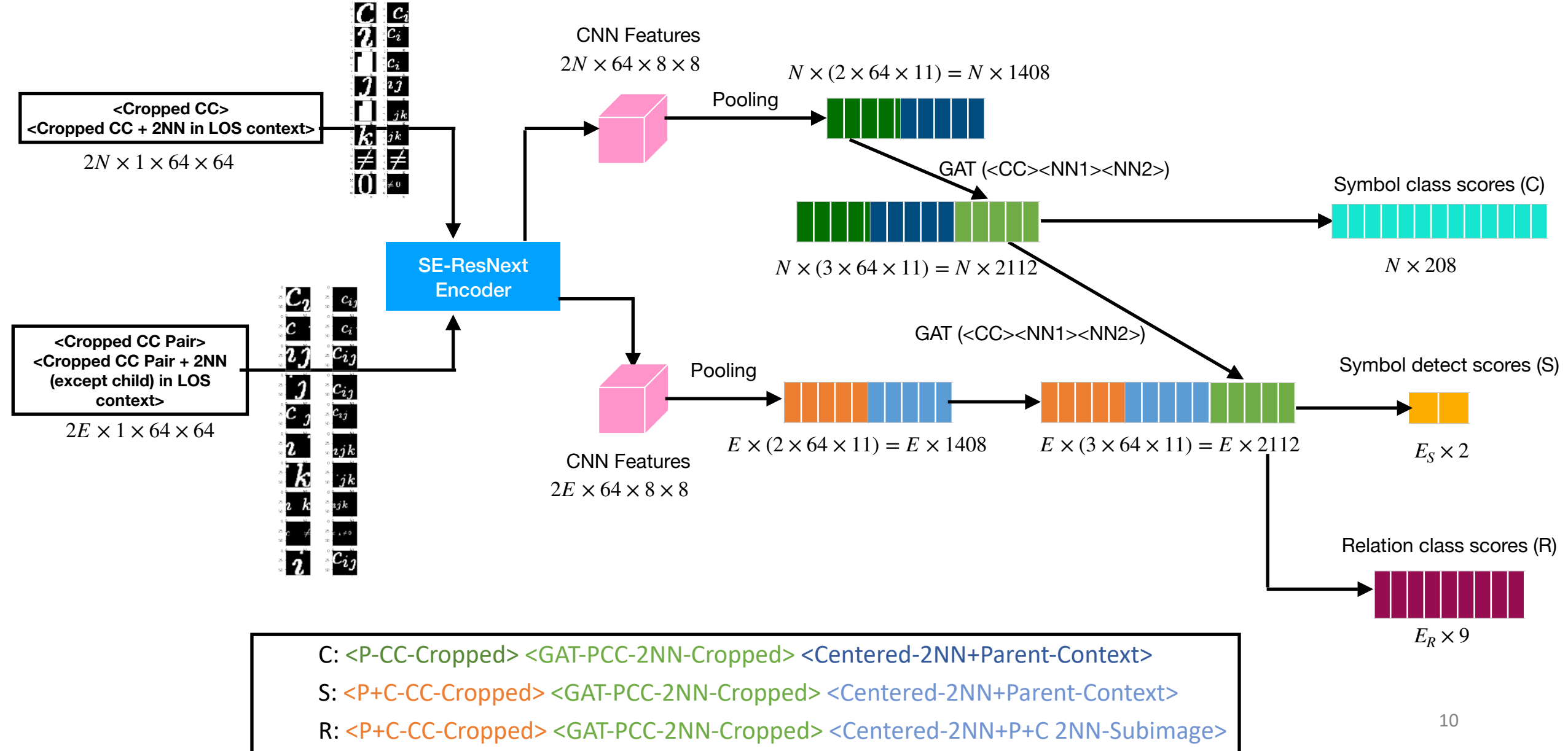
Spatial pyramidal pooling (SPP) Average - 3 levels to replace average pooling

Reduced encoder output channels from 512 to 64

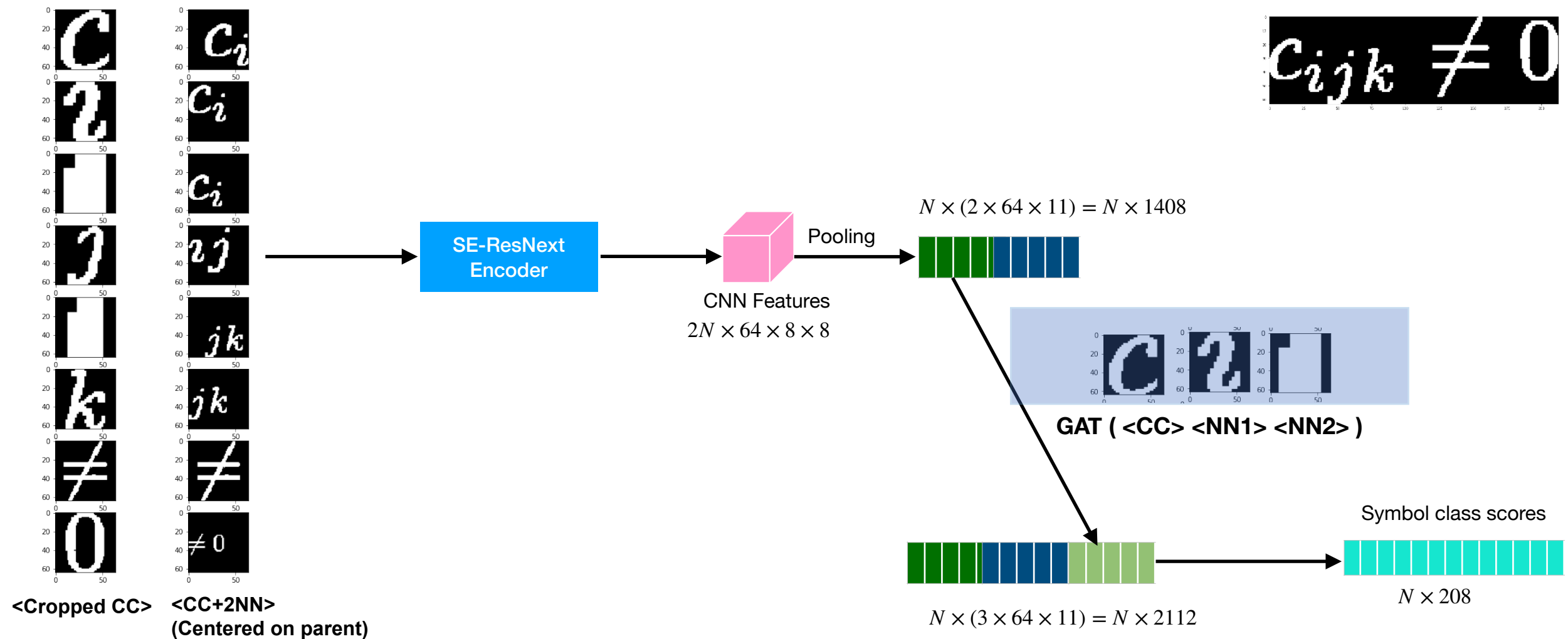
Feature length: $64 \times 11 = 704$ vs. 512



CNN Design: Shared features



More Context: Graph Attention Network (GAT)

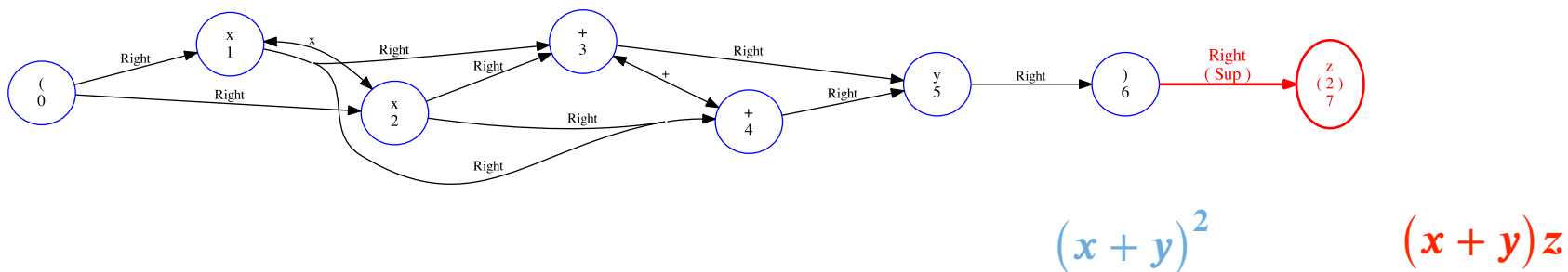


Results: Typeset Formula Images

Training: InftyMCCDB-2¹ train -> 12551 formula images

Testing: InftyMCCDB-2¹ test -> 6830 formula images

	Feature vector length	Symbols		Relation		Structure		Execution time (ms per formula)
		Detection (F1)	Det. + Class (F1)	Detection (F1)	Det. + Class (F1)	Detection (% file)	Det. + Class (% file)	
[1] QD-GGA	C, S, R: 2 x 704	98.32	95.66	94.85	94.35	89.27	83.27	363.54
LGAP	C, S, R: 3 x 704	99.48	98.6	97.11	96.61	91.67	88.11	98.4



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

[1] Mahdavi, M.; Sun, L.; Zanibbi, R.: **Visual Parsing with Query-Driven Global Graph Attention (QD-GGA)**. In Conference on Computer Vision and Pattern Recognition Workshops (2020)

Results: Handwritten Formulas

Training: CROHME 2023 Train: 10,979 inkml

Test dataset: CROHME 2019 Test: 1199 inkml

Submitted to CROHME 2023 Competition²

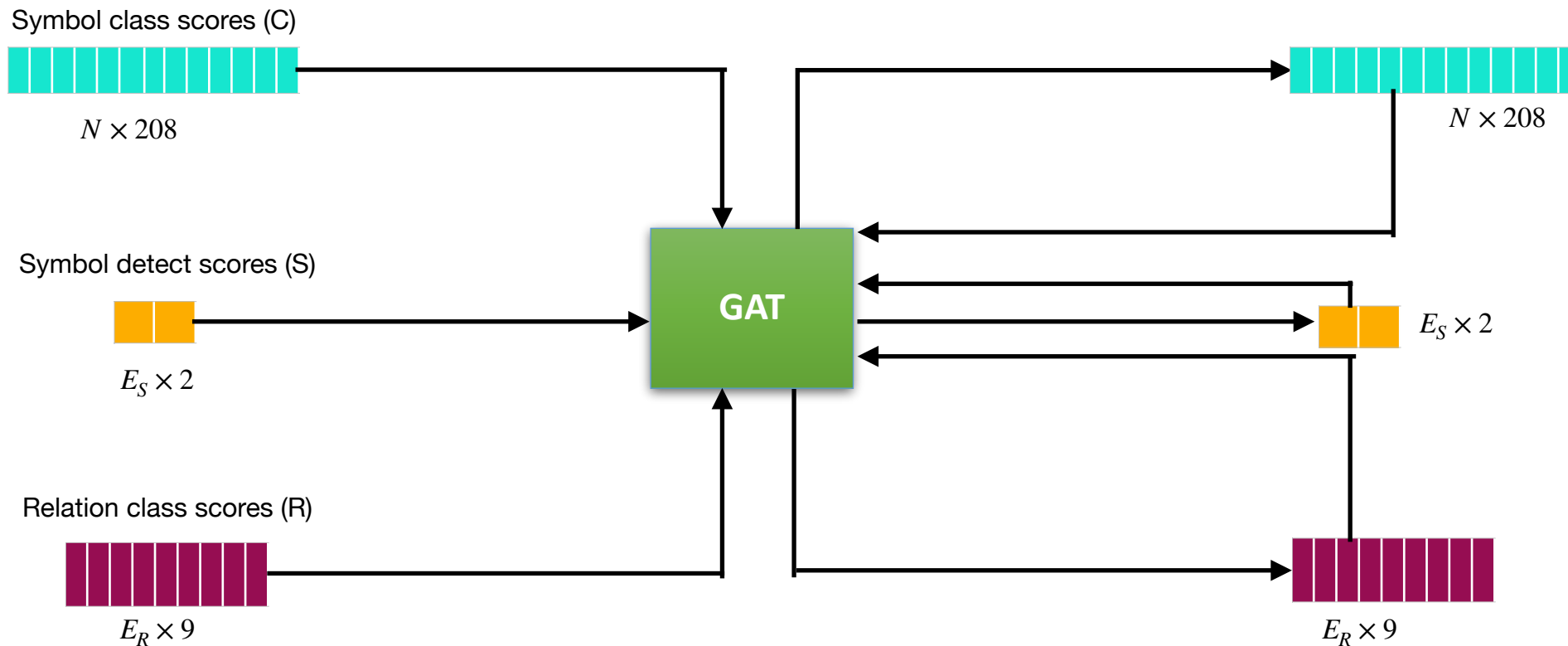
	Feature vector length	Symbols		Relation		Structure		Execution time (ms per formula)
		Detection (F1)	Det. + Class (F1)	Detection (F1)	Det. + Class (F1)	Detection (% file)	Det. + Class (% file)	
[1] QD-GGA	C, S, R: 2 x 704	94.35	82.08	86.8	85.16	57.3	25.6	1000.83
LGAP	C, S, R: 3 x 704	94.67	86.44	79.55	77.15	49.46	30.11	216.01

² <https://crohme2023.ltu-ai.dev>

[1] Mahdavi, M.; Sun, L.; Zanibbi, R.: **Visual Parsing with Query-Driven Global Graph Attention (QD-GGA)**. In Conference on Computer Vision and Pattern Recognition Workshops (2020)

Current Experiments

1. Training on formulas at primitive and symbol level together (shuffled)
2. **GAT-based task interaction** with recurrent message passing



GAT-based task interaction with recurrent message passing

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 interaction between the three tasks using GAT
- Match recognition rates with current SOTA models (encoder-decoder)
- Modify LGAP for parsing chemical diagrams

[2] Shah, A.; Zanibbi, R.: **Line-of-sight with Graph Attention Parser (LGAP) for Math Formulas**. in Document Analysis and Recognition – ICDAR 2023, in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2023, **Submitted**

[3] Code: <https://gitlab.com/dprl/qdggg-a-parser>

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, B. Amador, and M. Langsenkamp for their contributions to the formula extraction, parser and search.



Alfred P. Sloan
FOUNDATION



Q&A

Motivation: Inf. Retrieval - Question Answering

What is the value of
this integral?

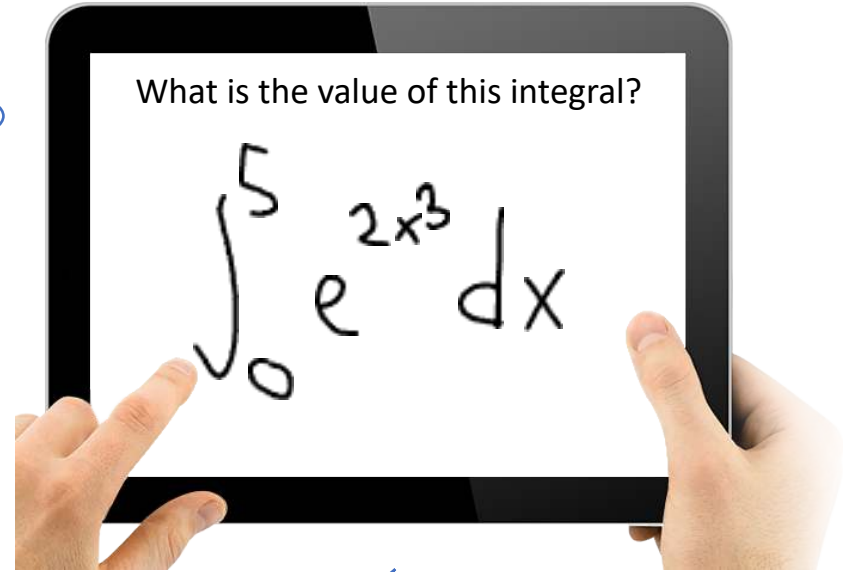
$$\int_0^5 e^{2x^3} dx$$

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?



LGAP Error Analysis on CROHME

Symbols: z (2), X (x), alpha (a), C(LeftPar), a_1 (n), 6(b), 1(|)

Relations:

- Missed Above Relations - 42 errors (most cases caused by merging fraction and 1 as 1) for half written 1 above fraction)

$$x = \frac{2\pi}{\sqrt{2}} \left(m + \frac{1}{2} \right) \left(\frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \right)$$

$$a + \frac{1}{2} \geq \frac{1}{2}$$

$$\sqrt{an} \ln f^{acb}$$

$$M_2 \left[I(I+1) - \frac{1}{4} Y^2 \right]$$

PUNC

