

LGATI Overview

A visual parser which produces an output SLT from math formula images or handwritten strokes

1. Create Line-Of-Sight (LOS) graph over primitives (strokes/CCs)
2. Score segmentation, symbol, and relationship hypotheses using a multi-task CNN
3. Segment and classify symbols
4. Select relationships using a maximum spanning tree with Edmond's algorithm [1]

Advantages:

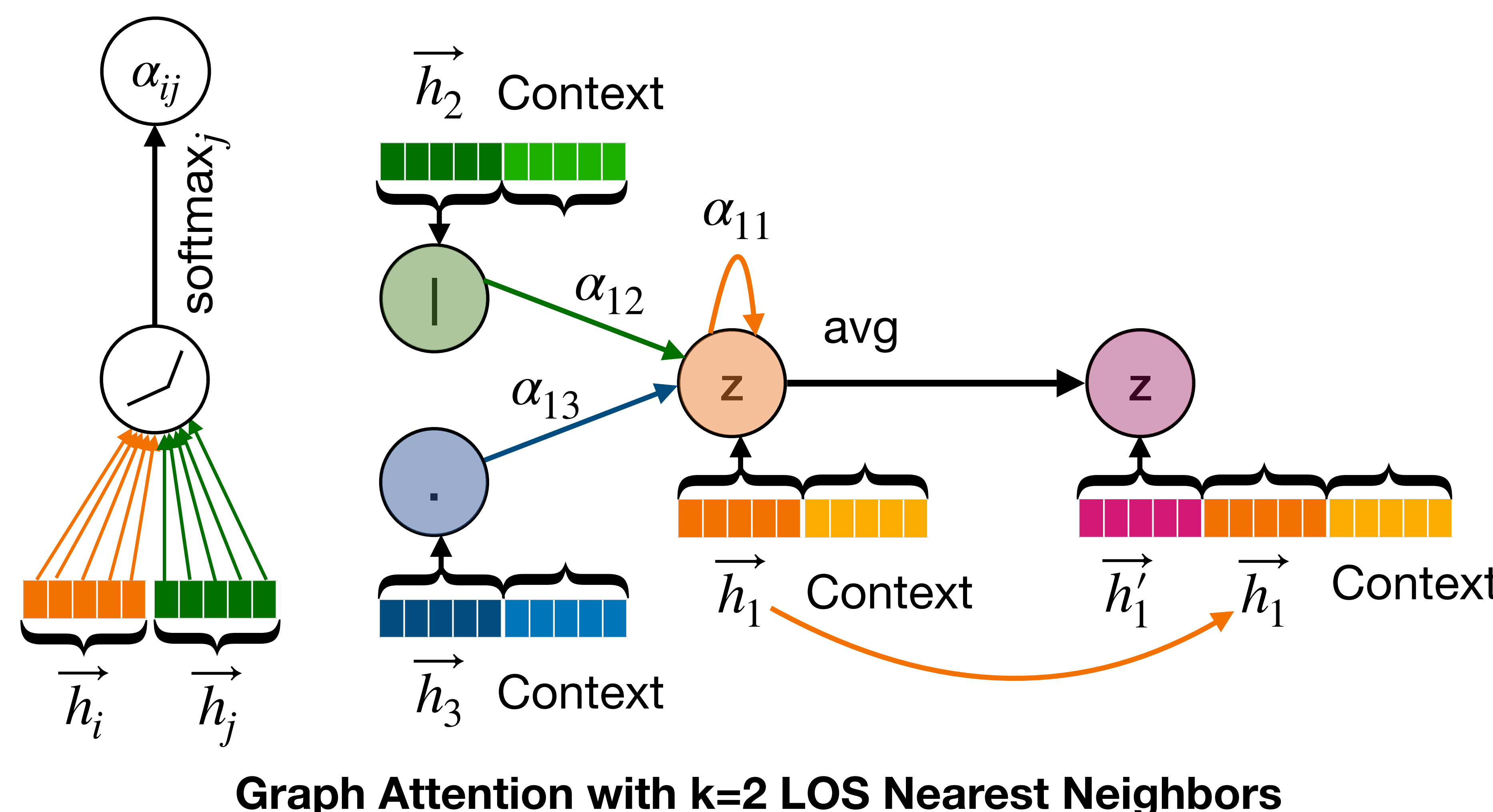
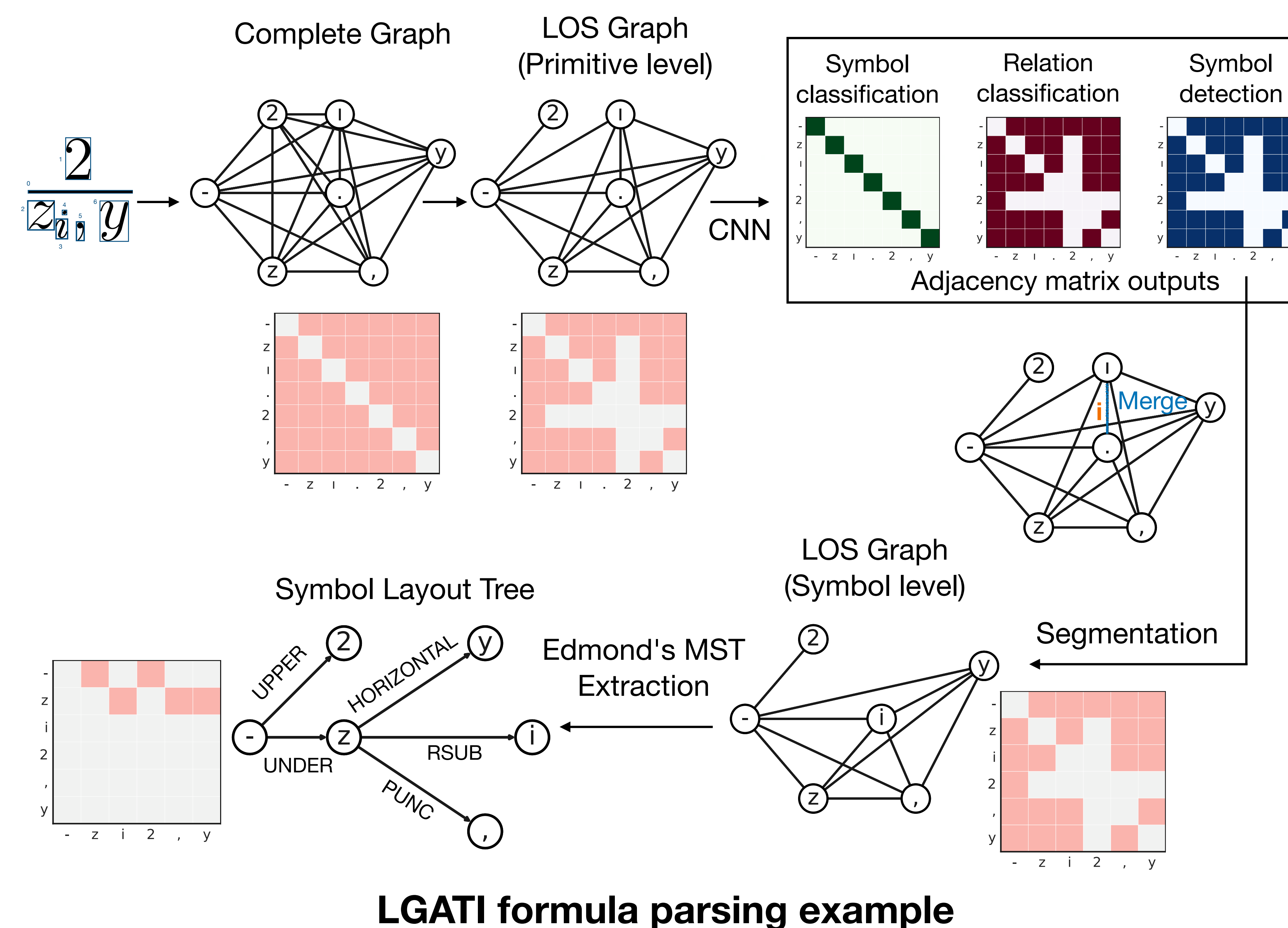
1. Enhanced error diagnosis at primitive level
2. Easy interpretability and accelerated inference
3. Mitigation of context limitations in previous graph-based models (e.g., QD-GGA [2])

Recognition Model Components

- **Inputs:** binary query and LOS attention masks
- **CNN:** builds upon QD-GGA [2], with SE-ResNext backbone and attention modules
- **1D context:** 1-by-3 convolution to capture contextual relationships
- **Spatial pyramidal pooling:** 3 levels (1, 2h, 2v, 3h, 3v)
- **Multi-Task Cross Entropy loss** with Adam optimizer

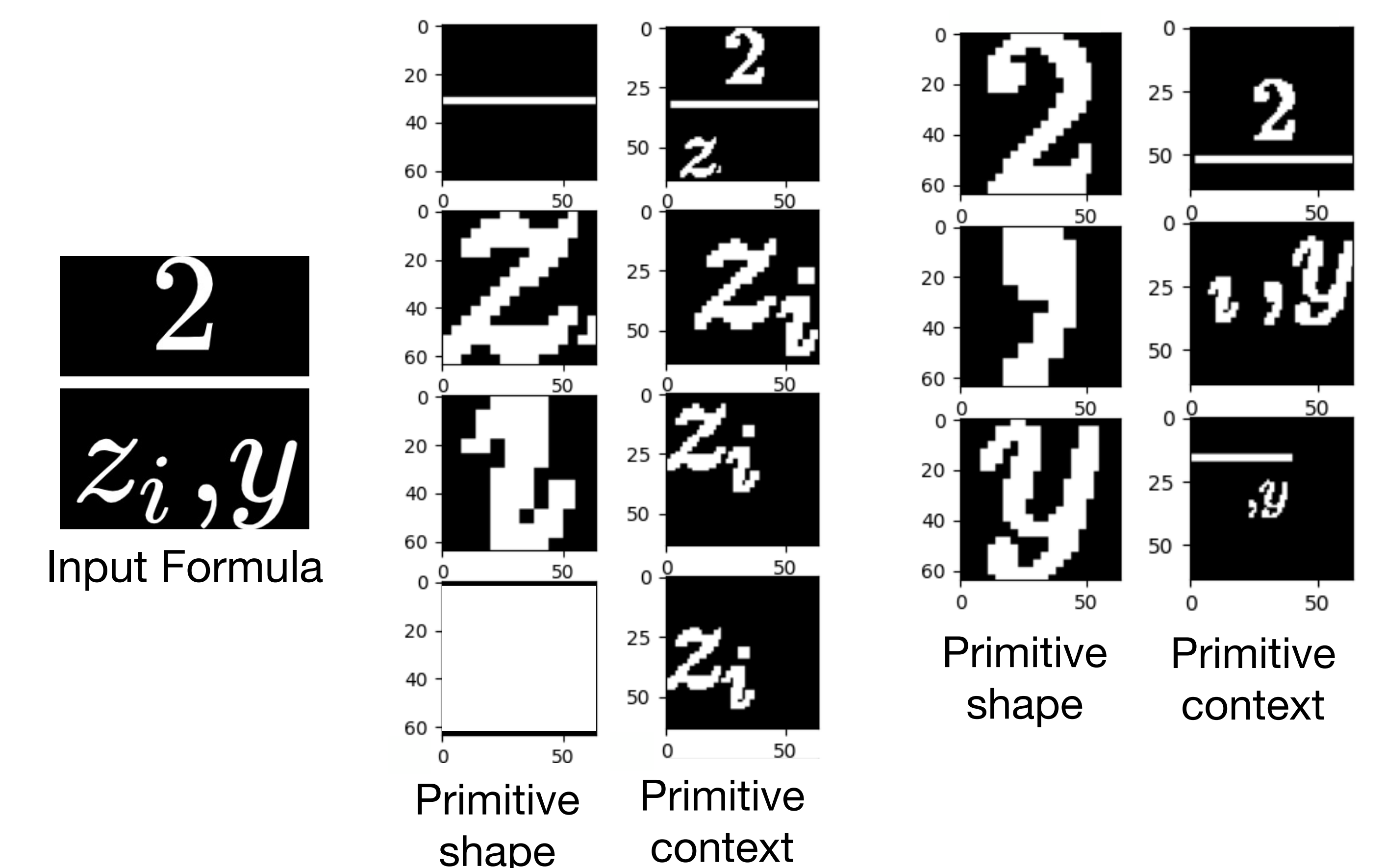
$$\delta(N, E) = \sum_{e=1}^{|E|} (CE(e, D) + CE(e, R)) + \sum_{n=1}^{|N|} CE(n, S)$$

- **Segmentation:** Merge primitives into symbols and average probabilities for merged primitives
- **Edmond's arborescence** to construct Symbol Layout Tree (SLT), avoiding duplicate edges with identical relationships



Graph Attention for interaction

- Modified GAT[4] with k-nearest LOS neighbors (k=2) for learning correlation between individual primitives
- GAT for interaction among 3 classification distributions
- 5 % improvement in formula recognition rate



Shape and Context Features

Other Contributions

- Consistent separation of shape and context visual features for all tasks
- Reduced number of segmentation and relationship hypotheses using LOS neighbors

Results

MST Model	Symbols		Relationships		Formulas	
	Detect.	+Class	Detect.	+Class	Structure	+Class
LGATI	99.48	98.60	97.11	96.61	91.67	88.11
LGAP	98.32	95.66	94.85	94.35	89.27	83.27
QD-GGA	98.50	94.54	94.43	93.96	87.77	76.72
LPGA _{RF}	99.34	98.51	97.83	97.56	93.81	90.06
LPGA _{CNN}	99.35	98.95	97.97	97.74	93.37	90.89

- Evaluated on 6,830 InftyMCCDB-2 test formula images (matrices and grids removed).
- LGATI takes 77 mins/epoch to train on 12,551 training images and 32 mins to process 6,830 test formula images (281 ms/formula)

Future work

1. Positional encodings for improved visual features
2. Extend parser for chemical diagrams