

## LGAP 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:

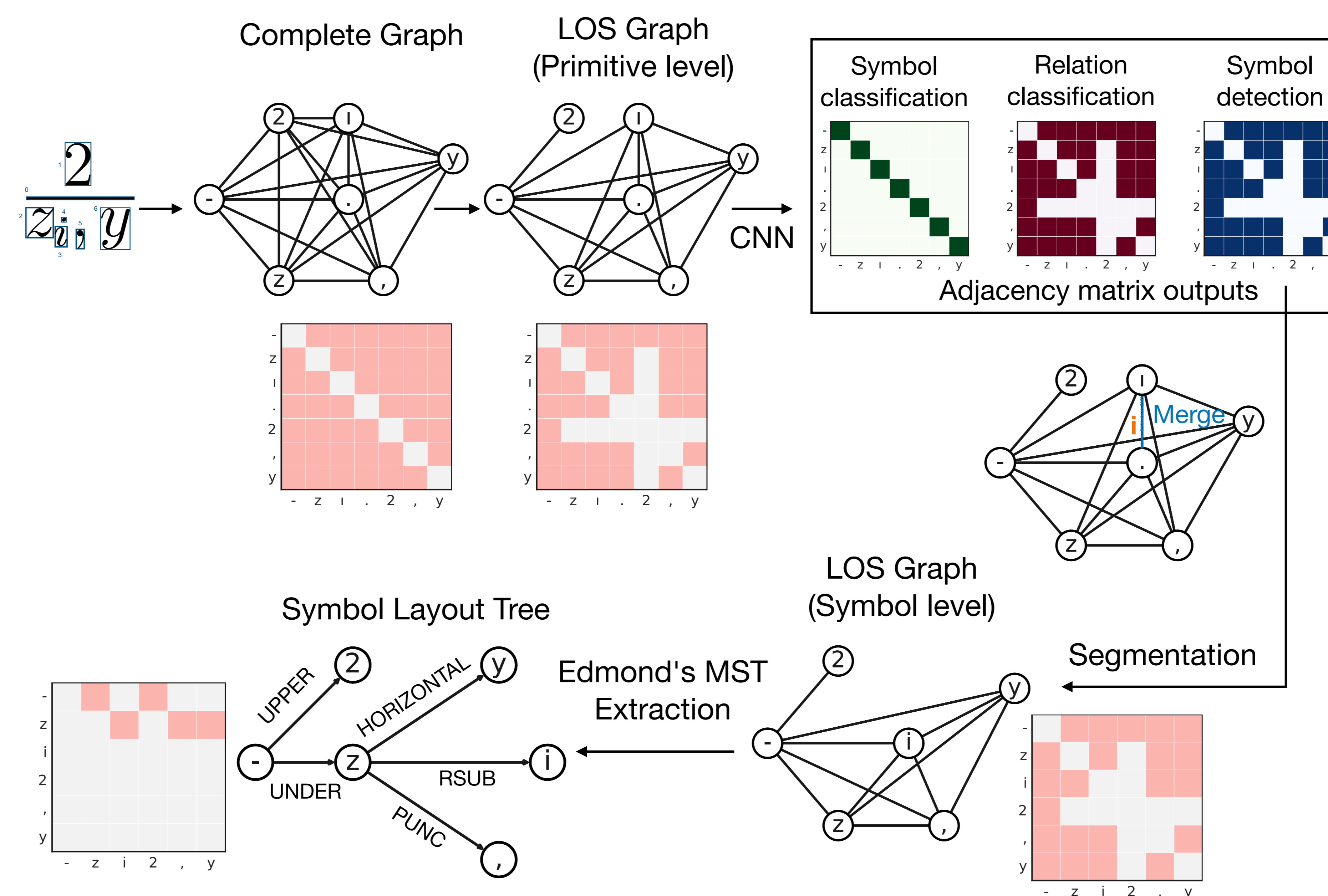
1. Enhanced error diagnosis at primitive level
2. Easy interpretability and accelerated inference
3. Mitigation of context limitations in previous graph-based models (e.g., QD-GGA [2])

## Recognition Model Components

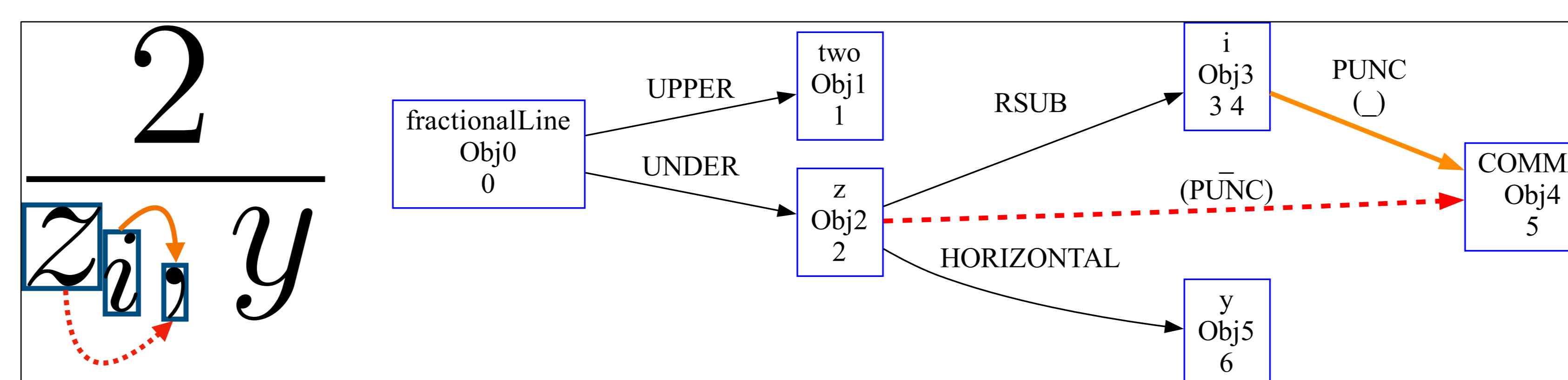
- **Inputs:** binary query and LOS attention masks
- **CNN:** builds upon QD-GGA [2], with SE-ResNext backbone and attention modules
- **1D context:** 1-by-3 convolution to capture contextual relationships
- **Spatial pyramidal pooling:** 3 levels (1, 2h, 2v, 3h, 3v)
- **Multi-Task Cross Entropy loss**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



LGAP formula parsing example



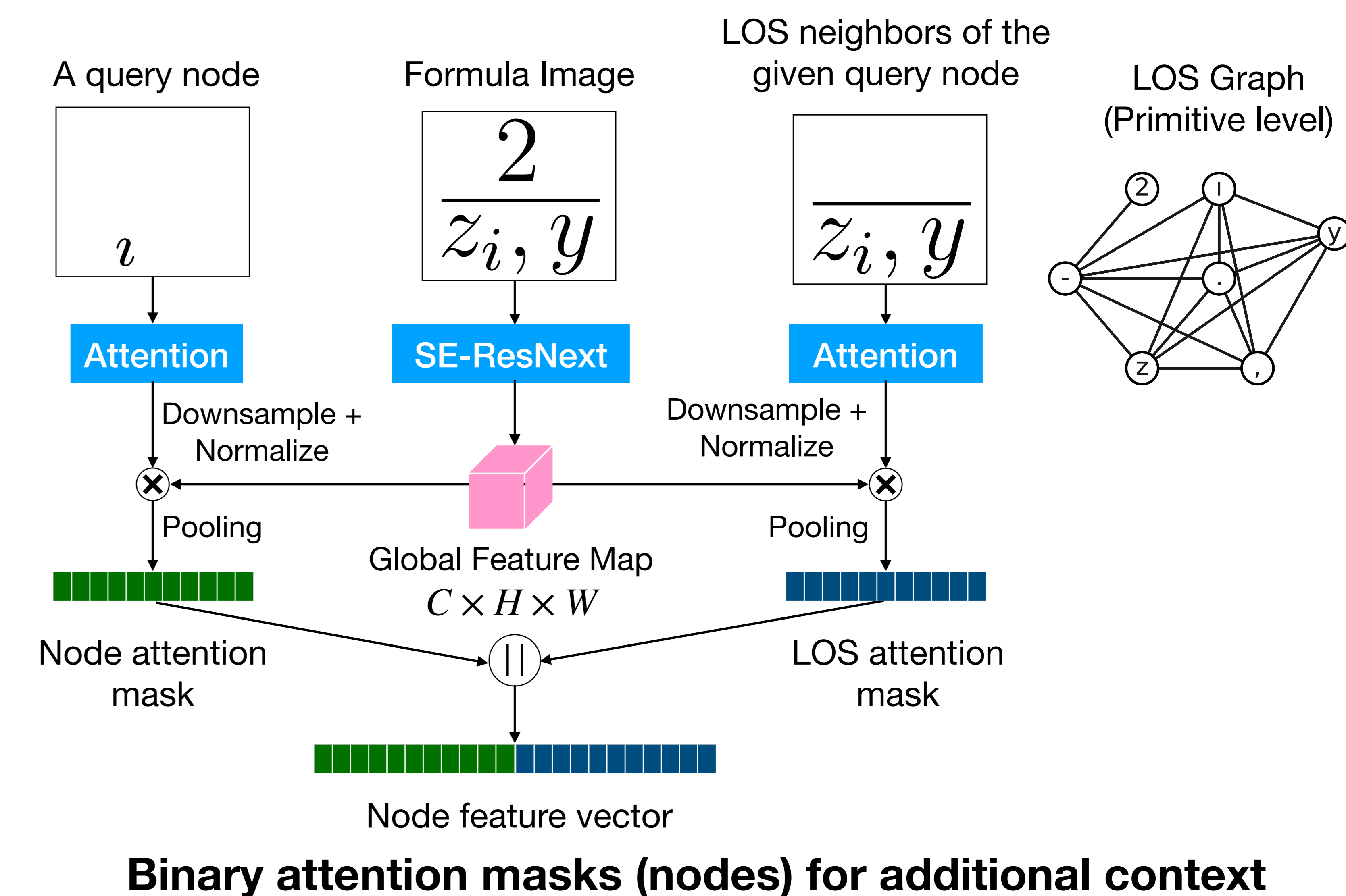
Modified punctuation (PUNC) ground truth representation

## Attention and Context Enhancement

- Query binary masks focus CNN on relevant image regions for specific task
- Additional LOS binary masks provide contextual spatial information from LOS graph neighbors
- Task-specific convolutional blocks (three kernels) act as attention modules for nodes, edges, and LOS masks

## Contributions

- Improves QD-GGA [2] accuracy while preserving interpretability
- Represent punctuation relationships more consistently
- Additional visual context from LOS graph neighbors
- Spatial pyramidal pooling rather than single-region average pooling of convolutional features



## Results

MST Model	Symbols		Relationships		Formulas	
	Detect.	+Class	Detect.	+Class	Structure	+Class
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 <sub>RF</sub>	99.34	98.51	97.83	97.56	93.81	90.06
LPGA <sub>CNN</sub>	<b>99.35</b>	<b>98.95</b>	<b>97.97</b>	<b>97.74</b>	<b>93.37</b>	<b>90.89</b>

- Evaluated on 6,830 InftyMCCDB-2 test formula images (matrices and grids removed).
- LGAP takes 25 mins/epoch to train on 12,551 training images and 11 mins 12 secs to process 6,830 test formula images (98.4 ms/formula)

## Limitations and Future work

1. Extend evaluation to handwritten datasets
2. Explore GNNs to replace the sequential 1D module, which is based on spatial sorting order
3. Advance interactions among tasks
4. Extend parser for chemical diagrams