Roadmap Outline

Data Preparation

Refine on given training set

NIST – Refine (For English Characters)

CROHME (For math symbols): Online dataset to offline-like data : opencv.dilate + Gaussian blur

Model:

Classifier with following symbols:

Digits: 0-9

Characters: x, y, a, b, c, d, m, n, p, delta, f, h, k, sin, cos, tan, A, pi

Operators: +, -, \*, /, =, sqrt, ^, \_, bar, frac, cdots

Architecture: ResNet-50

Small window filters with various sizes to scan after opencv.findContours()

A Region Proposal-like Network

Merge decomposition parts of signs like =, /, cdots to one rectangle with RNN(LSTM)

From symbol locations (4 points) and its label to adjust label and write to data structure

Tree-like ADT to hold math expressions

From ADT to latex