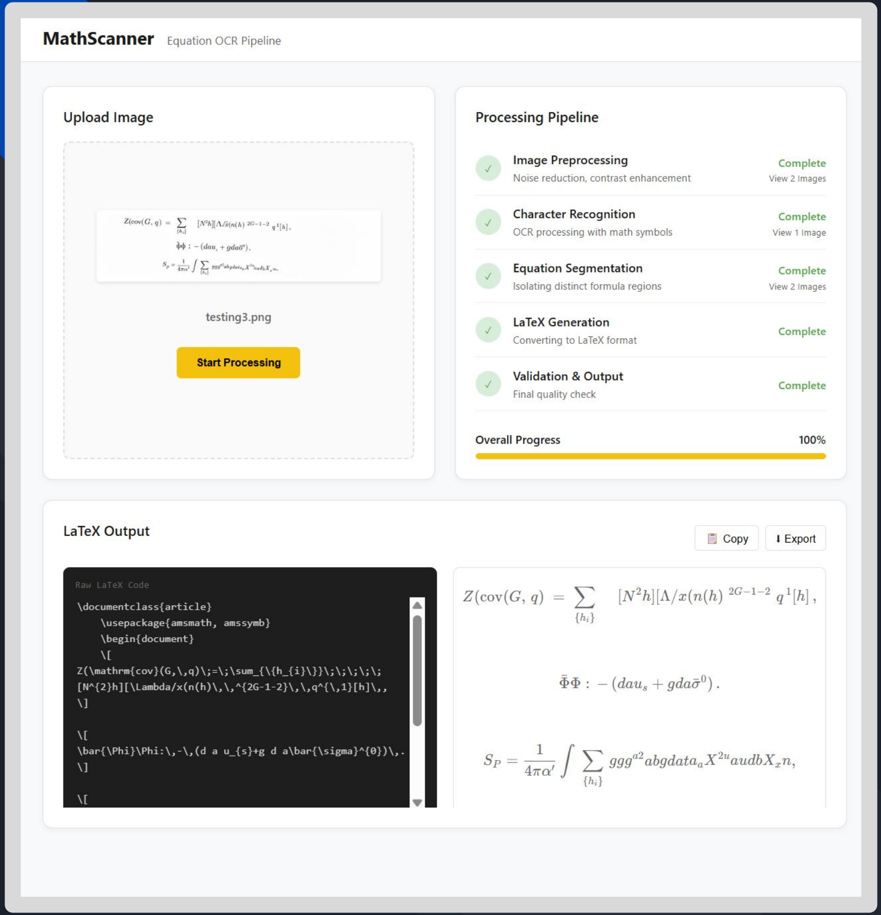
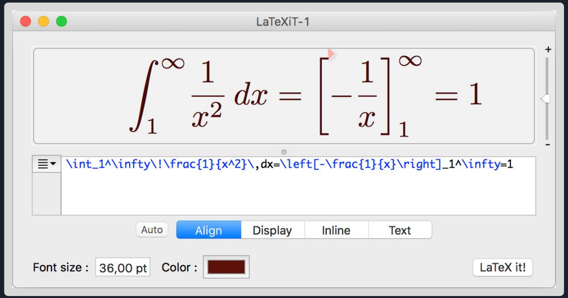


Paper2Proof

Archit and Bhaumik

Problem Statement



Motivation:

- Converting handwritten or printed mathematics into LaTeX is frequently required in technical fields but is time-consuming and tedious to type.

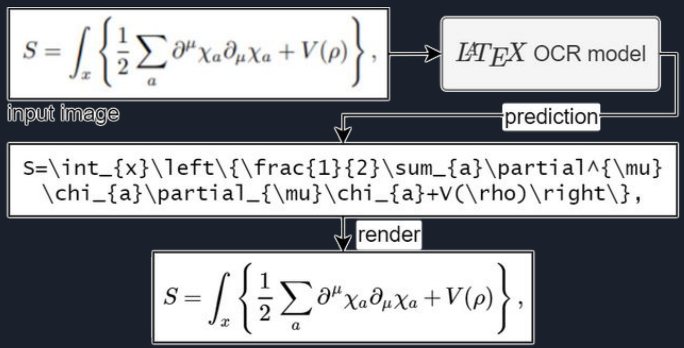
Historical Approach (Classical Computer Vision):

- Involved steps like denoising, character segmentation, inferring structural relationships with hand-engineered grammars, and reconstructing LaTeX.
- Very fragile: fail for overlapping strokes, handwriting irregularities, or touching symbols.

Modern Approach (Deep Learning):

- Sequence-to-sequence models (e.g., Im2LaTeX) and Vision Transformers (ViTs) like pix2tex have shown strong performance on single, clean mathematical formulas.

Our Approach



The Gap

- Lightweight, open-source deep learning models are good at single equations but fail on full pages
- Powerful end-to-end models (e.g., MathPix, GPT-4o mini) can handle full pages but are computationally heavy, require online inference, and are impractical for offline/low-resource settings.

Our Key Insight and Goal:

- We develop a **classical computer-vision pipeline** to isolate equations from full documents. This segmentation allows an existing lightweight model (pix2tex) to operate on full pages without retraining

More on Pix2Tex

Model	BLEU Score	Edit Distance	Token Accuracy
pix2tex (LaTeX-OCR)	0.88	0.10 (normalized)	0.60

Table 1: Reported performance metrics for the pix2tex (LaTeX-OCR) model

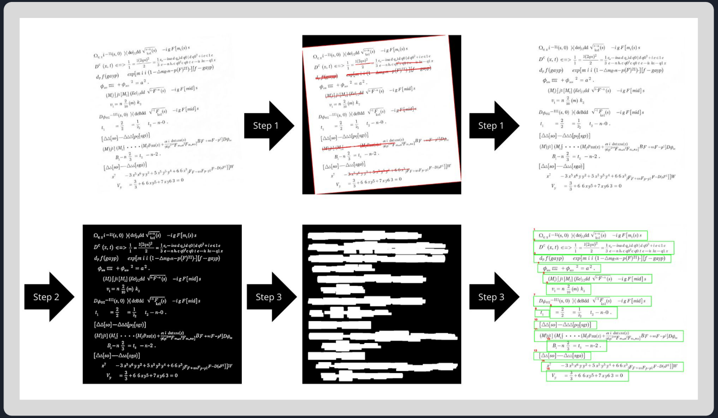
Background

- This model uses deep learning to generate the latex from a given input image of an equation
- It was trained on Im2LaTeX-100k and wikipedia equations
- It has a Vision Transformer (ViT) Encoder with a ResNet backbone and a Transformer Decoder to output LaTeX tokens

Limitations:

- The model exhibits a strong bias toward **Greek characters** because most of its training data consisted of them. It struggles to recognize digits, attempting to map them to their nearest Greek counterpart
- It shows difficulties with accents and confusion between visually similar symbols
- The model collapses and produces unusable output when multiple equations are merged into a single image, highlighting that it assumes the input contains exactly one formula

Design Choices - What doesn’t work



Skew Processing:

- PCA-based angle estimation and projection-profile methods were tested but proved sensitive to handwritten curvature and noise

Segmentation:

- **2D Morphological Dilation:** Rejected because it often merged adjacent equation lines, degrading segmentation precision, although it helped with handwritten superscripts.
- **Hierarchical Clustering:** Rejected as an alternative to overlap-based merging because it introduced sensitivity to hyperparameters and produced inconsistent grouping across handwriting styles

Live Demo - https://github.com/ArcCreate/Paper2Proof/