

Handwritten or Digital Notes? Why Not Both! Using Language and Knowledge Modeling to Improve OCR for Handwritten Class Notes

Angelo De Asis '26, Jason Farnsworth '26; Brad Spendlove, Computer Science

Hand-writing notes has demonstrable benefits for students, but many still choose to take notes digitally for convenience in storage, backup, and searching. Optical Character Recognition (OCR) is a technology for digitizing physical documents, but it may struggle to recognize and accurately transcribe hastily written and organized pages of notes. We propose incorporating contextual information relevant to class notetaking, such as topic analysis, language modeling, and factual knowledge, to improve

the efficacy and accuracy of an OCR

system for digitizing those notes. Starting with a state-of-the-art base-line OCR model, we will build new programs to incorporate that contextual information and measure its impact on the system's accuracy. We anticipate that our improved system could form the core of a digitizing app to help students combine the academic benefits of hand notetaking with the convenience of digital files.

