Research project
English handwriting recognition for exercises and exam papers in education sector

Objectives

- Deliverables will serve as a major product upgrade (subject coverage extension) for MachInk, an auto-marking system for primary and secondary schools, developed by Mach Innovation, an A.I. startup partnered with HKU SAAS Data Science Lab and secured offer of HK Science and Technology Park (HKSTP) Incu-Tech programme

Problem statements / Questions types to handle

Primary school general

- Fill in the blanks (case sensitive, short form, e.g. isn't/aren't)
- Tick the answer box / Circle the answer / Blacken the circle
- Short answers reply (capture key words and check grammar mistakes, misspelling)
- Proofreading (Circle the mistakes & write correct ones on the blanks)

Secondary school general

- Long answers reply (capture key words and check grammar mistakes)
- Proofreading (Circle the mistakes & write correct ones on the blanks / above)

Deliverables

- Built an English handwriting character recognition with extremely high accuracy (both single words and sentences after segmentation, case sensitive, punctuation)
- Improve current line segmentation algorithm to fit better the English exercises and exam papers
- Construct an algorithm that can capture key words within short/complete sentences

Skillsets requirements

- Solid programming skills in Python
- Rich experience in using machine learning libraries such as Tensorflow, Keras, Scikit-learn, etc.
- Excellent understanding of machine learning techniques and algorithms, especially convolutional and recurrent neural networks (CNN & RNN)
- Experience in character recognition and natural language processing (NLP) is an absolute advantage

Research project

Handwritten Mathematical expression recognition for exercises and exam papers in education sector

Objectives

- Deliverables will serve as a major product upgrade (subject coverage extension) for MachInk, an auto-marking system for primary schools, developed by Mach Innovation, an A.I. startup partnered with HKU SAAS Data Science Lab and secured offer of HK Science and Technology Park (HKSTP) Incu-Tech programme

Problem statements to handle

General Mathematics

- Segment multi-line expressions into individual single-line expressions
- Recognise basic 10 based numbers (including + and – prefix, decimal points)
- Recognise fractions, including mixed fractions, and numerator/denominator expressions
- Recognise basic operators (+, -, ×, ÷, /)
- Recognise different level of brackets
- Recognise equal sign (=), at the front of the expression or in the middle
- Recognise percentage sign (%)
- Recognise common constant symbols (e.g.: π)
- Recognise common English alphabet variable symbols (e.g.: a, b, w, x, y, z, n, …)
- Recognise superscript features (e.g.: x^2, 14^{4+5}, 2^x)
- Recognise crossed out terms (e.g.: $2+\cancel{x}+y = 4+\cancel{x}$)
- Ignore scanned noise

Primary school specific features

- Tolerance of abnormal spacing due to bad writing or erasing of terms
- Recognise time representation (e.g.: 9:45)
- Recognise basic units (e.g.: °, mm, cm, m, kg, g, s, hr)
- Segment Chinese or unknown units

Deliverables

- Improve current multi-line segmentation algorithm to fit better then Mathematics exercises and exam papers
- Develop a notation / data structure or choose an existing one for the recognised Mathematical expression
- Build a handwritten Mathematical expression recognition with high accuracy (output top possibilities sorted by confidence level)

Skillsets requirements

- Solid programming skills in Python
- Rich experience in using machine learning libraries such as Tensorflow, Keras, Scikit-learn, etc.
- Excellent understanding of machine learning techniques and algorithms, especially convolutional and recurrent neural networks (CNN & RNN)
- Experience in handwritten numbers and character recognition is an absolute advantage