DESCRIPTION OF THE METHOD

Name: TAQ system

Author: Nam Quan Nguyen

Participant: Nam Quan Nguyen[a], Tran Hai Anh Vo[a], Quoc Thang Nguyen[a].

Address: (a) Cinnamon AI Lab Inc.

Our method includes principles of operation and steps:

1. **Text and Non-text classification**
   In this step, we use an encoder-decoder model deep learning with an variant architecture of U-net [2] and training with multi-task (7 tasks) to segmentation each pixel belong to text or non-text class. Specifically, 7 tasks are text mask, non-text mask, non-text contour mask, 4 sides of text component (top, bot, left, right).
2. **Smooth text components**

   First, we extract text components from text mask in step 1.

   Then we binarize image, extract connected components (CCs) and based on [1], we classify each CC belongs text or non-text class.

   Finally, we extract all text line with textual CCs and only keep strong text lines to smoothing text mask from step 1. A text line is strong if have some textual CCs (>= 3) and homogeneous about width, height of textual CCs.
3. **Non-text classification**

   First, we extract non-text components from step 1 and use non-text contour mask to eliminate redundancy components

   Then get all text components is inside of each non-text components and classify to table, image, or chart class.

Reference:

[1] T. A. Tran, I. S. Na, S. H. Kim, “A Robust System for Document Layout Analysis using Multilevel Homogeneity Structure,” Expert Systems With Applications, vol. 85, pp. 99-113, 2017.

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015