Deep Learning for Recognition of Japanese Historical Documents

LE DUC ANH　　データ科学研究系　特任助教

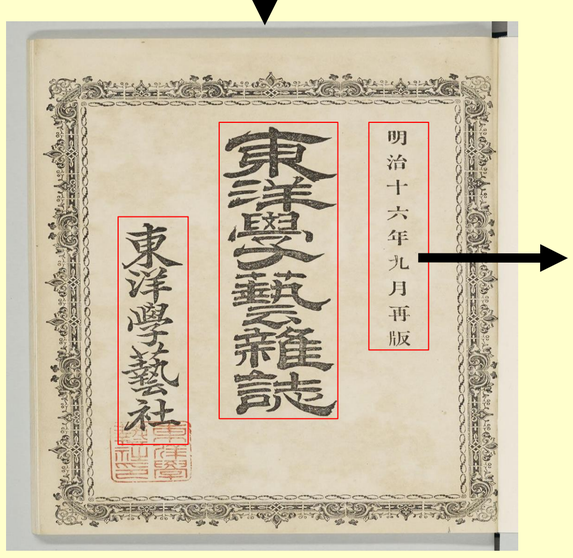
Introduction

- CODH has the following missions toward the promotion of data-driven research and formation of the collaborative center in humanities research.
- Employ state-of-the-art technologies in AI and Deep Learning to digitize Japanese historical document
- Construct data platforms for sharing historical document, digitization methods to promote trans-disciplinary participation of people with diverse backgrounds
- Aggregate, process and deliver humanities knowledge from Japan to the world through collaboration across organizations and countries

System Overview

Input image

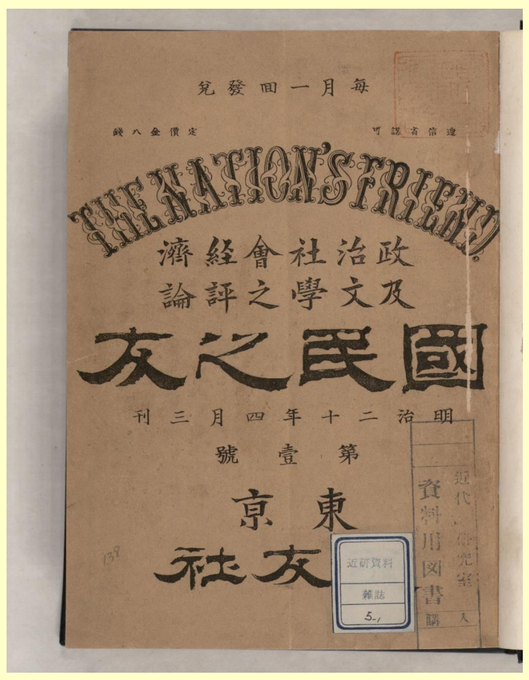
Layout analysis
paragraph, line, word

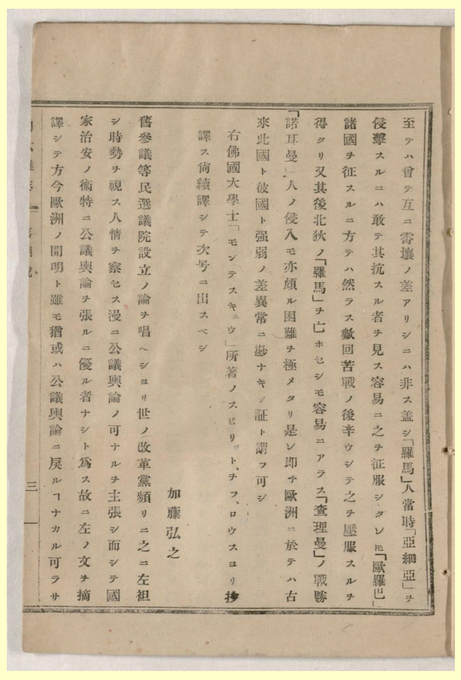


明治十六年九月再版

Text Recognition

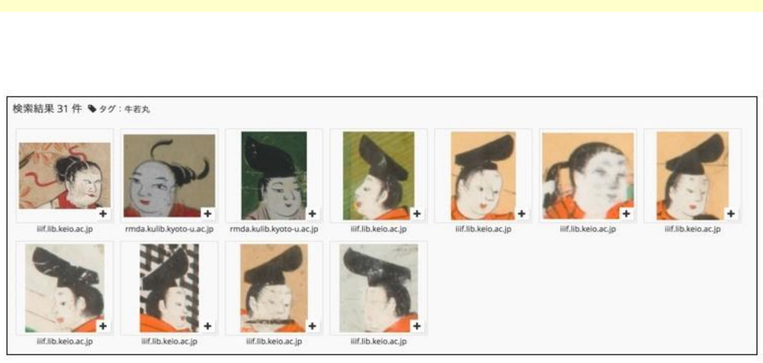
Database





Japanese Modern Magazines (~1900)

Pre-Modern Japanese Text



Collection of Facial Expressions from Japanese Emaki

System Architecture

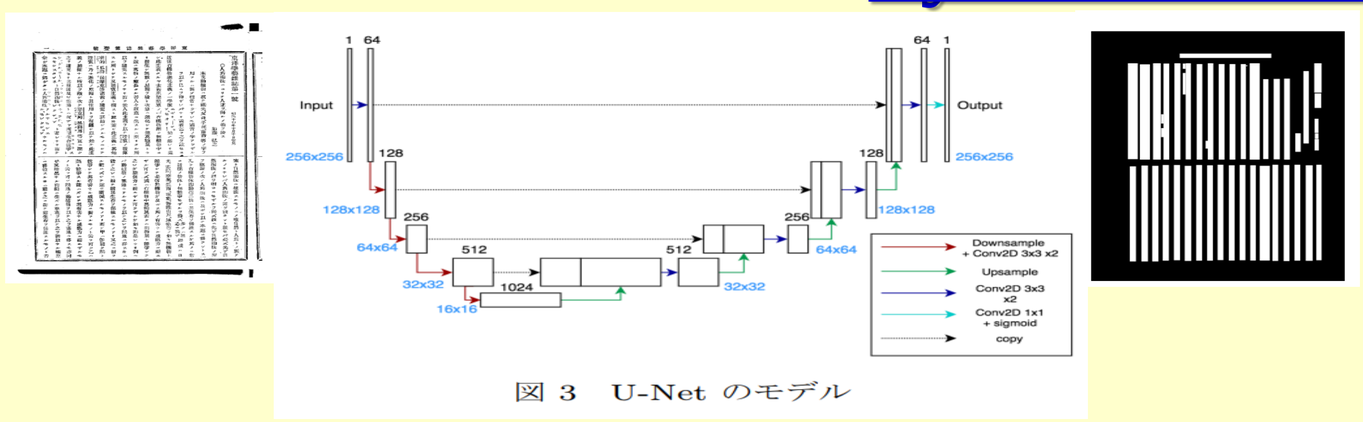
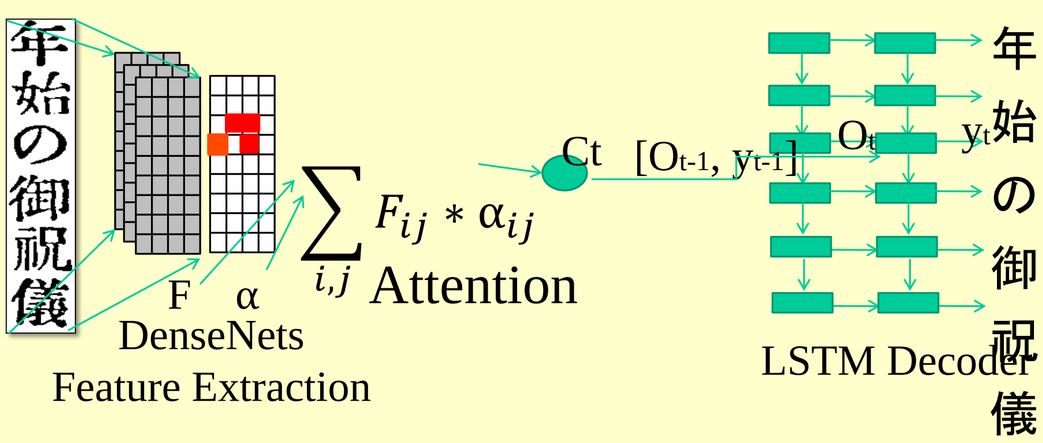


図 3　U-Net のモデル

Text Detection



Text Recognition