Practical machine learning and deep learning course project deliverable 1

Automated manga cleaning using CNNs

Team:

Vyacheslav Sergeev
vy.sergeev@innopolis.university

Nikita Strygin
n.strygin@innopolis.university

Dinislam Gabitov
d.gabitov@innopolis.university

Summary of project focused efforts

During the last two weeks we have decided to focus our efforts mostly on obtaining the data necessary for the project. The data we require is manga and somehow separated/marked text from the pages of said manga. The following paragraphs detail the data we have found and the challenges we have experienced.

Manga109

One of the potential sources of data we could use was dataset manga109 located at http://www.manga109.org/en/ . The dataset compiles several rather old manga sources and is intended only for scientific use. The access to dataset is provided on a case-by-case basis after filling out the application form. We have filled said form and gained access to the dataset.

We have identified some problems that exist in the manga109 dataset:
1. Data is rather old. Old manga has a drastically different artstyle from the newer manga. In addition to that recent manga artists experiment more with the position and style of the text that authors of some older manga. All of this means that if we train our solution only on manga109 we may experience generalization problems.

Example of older manga (on the left) and more modern manga (on the right):

2. The quality of data is dar from perfect. Some of the text on the pages is not marked. This quality issue can definitely lead to problems with neural network, so if we want to use the data we will have to hand pick it.
An example of quality issues (2 speech bubbles are not annotated):
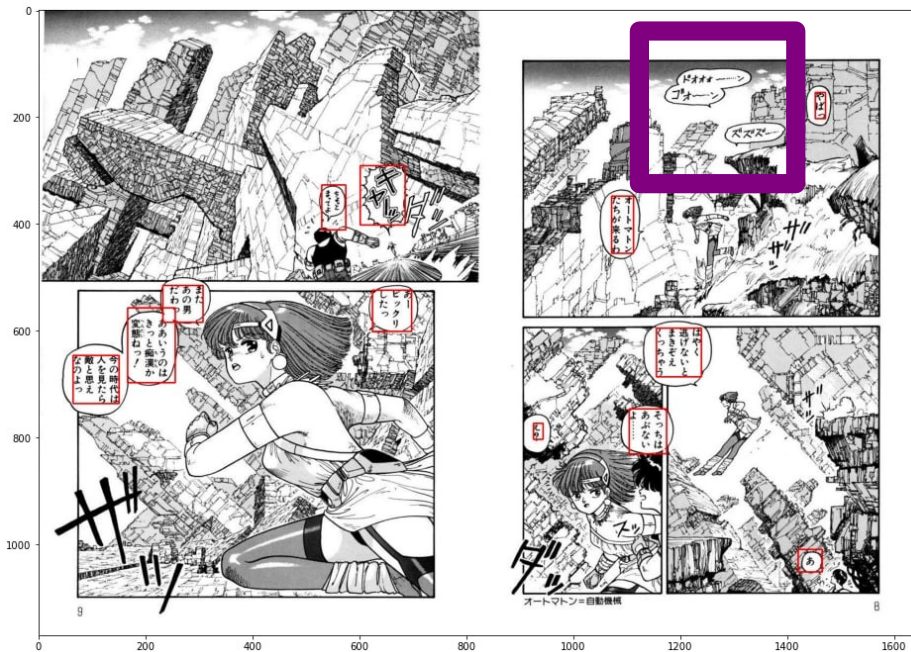

Trying to collect our own dataset

With manga109 being a bit too old and with its quality issues we wanted to try collecting our own dataset. However collecting it has proven to be a challenge for the following reasons:
1. Some of the manga has textured background. Which makes erasing text from it complicated. Normally you could paint the text over with white color, however with textured background even simplest speech bubbles become non-trivial.

Example of a textured background:
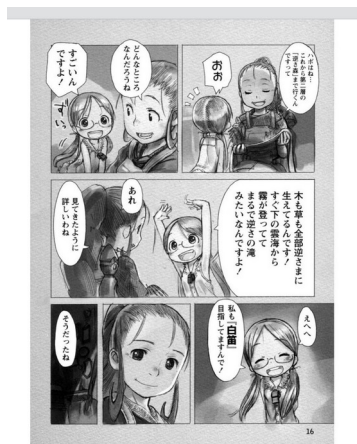
2. Some manga containt written sound effects which are nearly impossible to separate from the the images. But leaving them in means that neural network can get confused about the text it needs to find because the sound effects are essentially stylised text. Which leaves us with the only option – to ignore the pages that include them

Example of a manga page with a lot of sound effects:

3. Sometimes the ordinary text ends up obscuring part of the images which makes separating them a very difficult task. We decided to ignore the most difficult to annotate images as otherwise that would mean doing a lot of work.
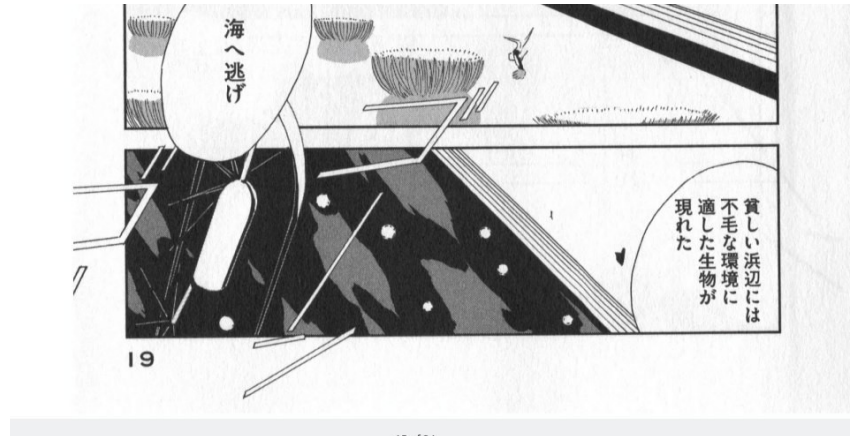An example of translucent speech bubble:

4. Miscellaneous things. Sometimes manga artists draw all sorts of things that make cleaning manga difficult.

Examples of weird things that can be found in manga:

Stylized text out partially out of bubble

Sound effects span multiple panels

Unconstrained Text Detection in Manga Dataset

After trying to come up with our own dataset we realised that our job would be a lot easier if only someone would have collected a dataset before us. After closer inspection it turned out that in one of the papers the authors have cleaned up manga109 and separated difficult and easy examples for us.
Example of the dataset annotations:

Source:https://zenodo.org/record/4511796#.YzsjCHbP1PY

Model architecture

Meanwhile we decided to start experimenting a bit with models. Currently we follow the paper from here: https://www.arxiv-vanity.com/papers/2009.04042/ since the authors claim to have achieved a nice result. We are currently trying pre-trained resnet34 + Unet. Model architecture summarized in an image:
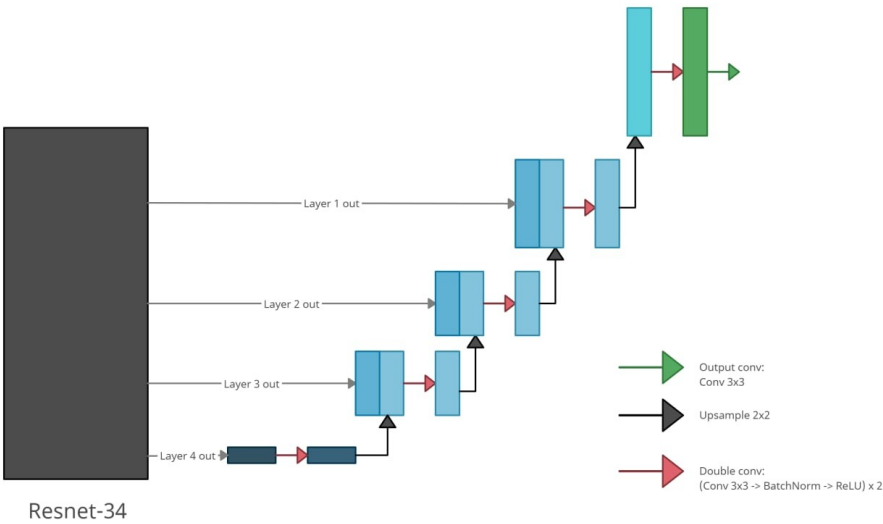
The results that we have managed to achieve so far:

Prediction   Ground truth   Input
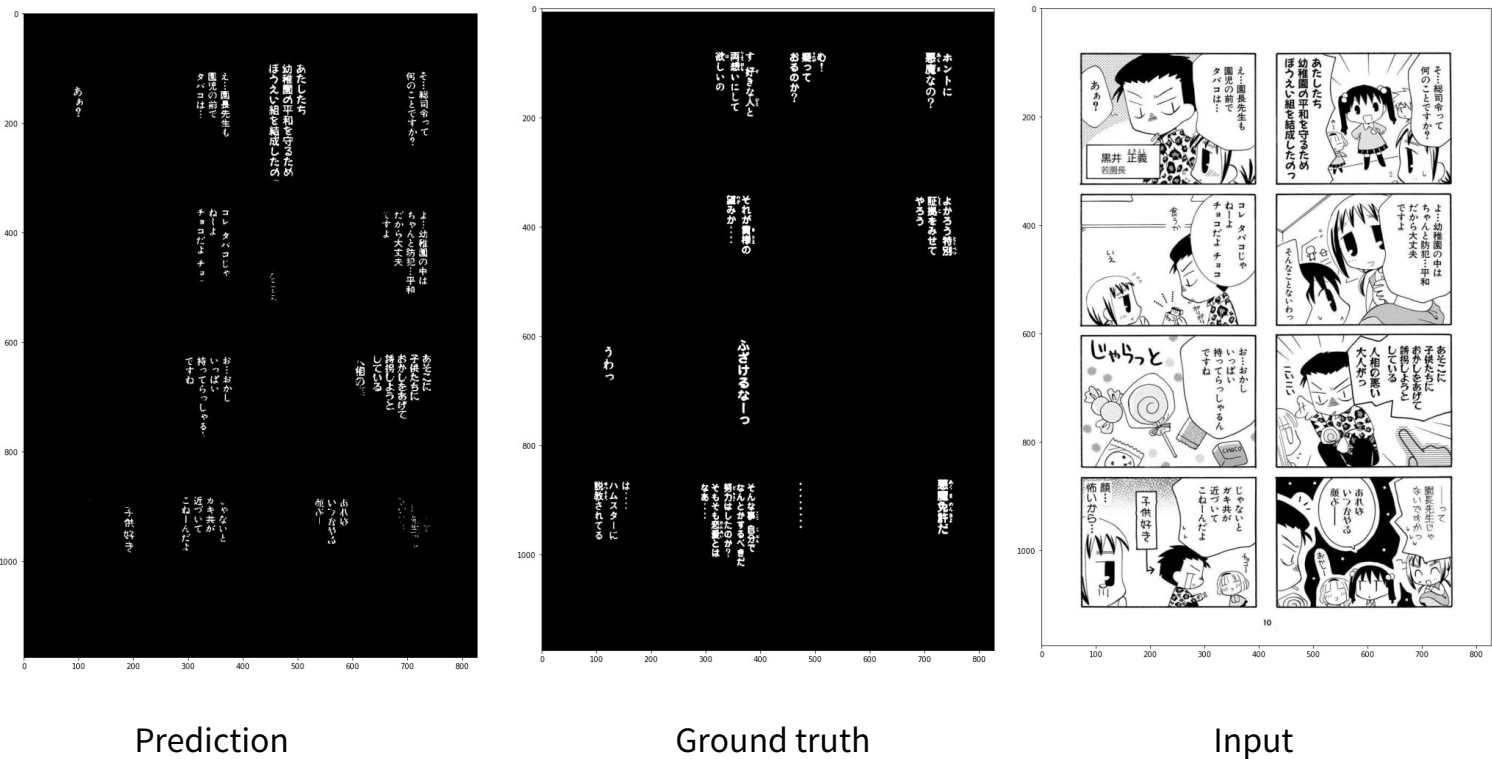
We still have ways to go, at least in terms of training the model. However the result already pretty closely resembles the desired output.

Direction of future work

We imagine the direction of the future work on the project as follows:
1. Clean up the dataset. (Mark all text or remove unmarked text)
2. Try cutting manga into pieces to learn model on pieces (will help avoid bad data)
3. Finetune resnet34 for our task
4. After steps 1-3 check the results on some recent manga
5. Add more items into the dataset (by hand or find some new datasets)
6. Pray to god that the model will generalise enough on the manga