

Practical machine learning and deep learning course project deliverable 2

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

The last few weeks definitely were not the easiest. We had exams and still have some major and difficult assignments. Honestly speaking, we do not have much to report this time. We, however, have produced a github repo with the code that we have so far and have continued our search for data.

The repository is located here: https://github.com/Dinislam36/Auto_manga_cleaner

However we cannot redistribute the data that we have collected due to Manga109 being closed-access and other manga being copyrighted. What is worse, the data in total weights quite a bit (~5Gb compressed) and is hard to fit & use in the cloud.

Member Contribution

We are currently satisfied with how group work has progressed ~~(when it did)~~.

Nikita Strygin: Data search, Data cleanup

Dinislam Gabitov: Model construction, Model learning

Vyacheslav Sergeev: Model research, Project Management

What the future (probably) holds

Week 1&2:

- * Cleaning Manga109 data from every not annotated text
- * Annotate additional data from downloaded manga
- * Learn the model on newly cleaned data
- * Evaluate the performance of the model
- * Search for possibly better models

Week 3&4+:

- * Try dataset augmentation with different fonts, text positions etc.
- * Construct dataset for inpainting
- * Try learning the model for inpainting
- * Try constructing photoshop plugin