Hungarian Helper

Indra Narayan Dutta(NSN8A7)

Objectives

- Create a model that acts as an OCR (Optical Character Recognition) for Hungarian characters – the OCR will be used for translation apps

- A lot of work has been done for English natural scene character recognition but hardly any models are available for Hungarian text

Challenges

- **Dataset Unavailability**: Unlike English, not a lot of datasets are available for Hungarian natural scene characters

- **Model size:** An OCR forms the last step in any scene text recognition pipeline and must therefore be light enough to allow bigger prior model pipelines
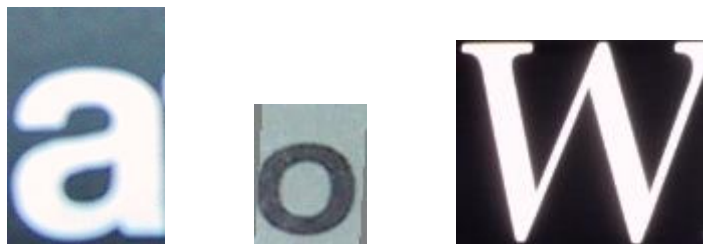
Dataset Preparation

- How the dataset challenge was solved:

  Since Hungarian and English share the same Latin script and only differ in the Hungarian special characters, I combined pre-existing English scene text images with Hungarian Characters from a different domain

  English Scene Text from ICDAR Challenges

  Hungarian special characters from handwritten text recognition
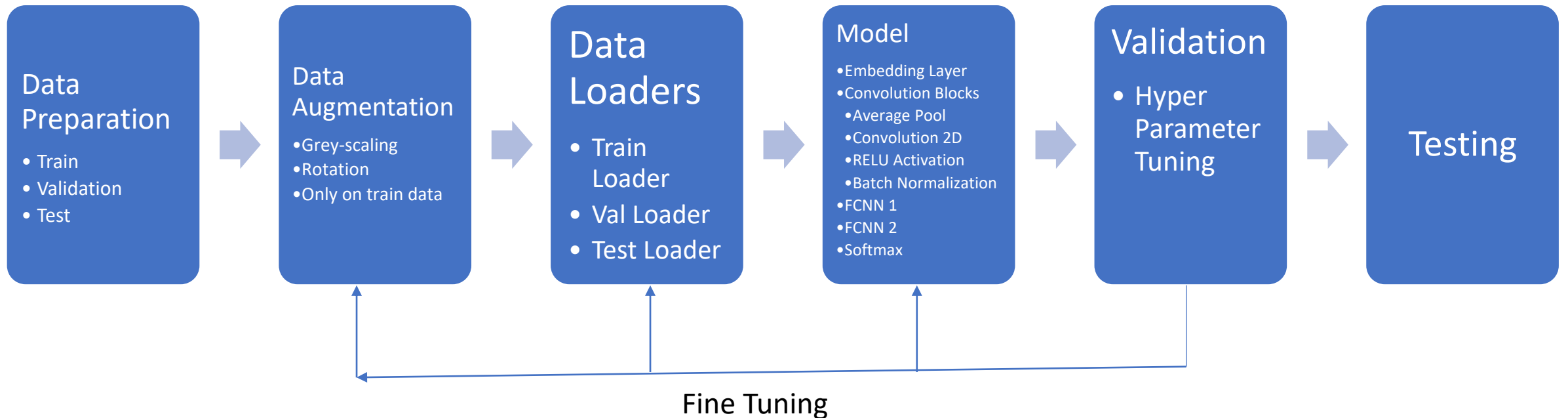

Dataset Split

- The following datasets were considered:
  - ICDAR 2003 Challenge for English Natural Scene Text Images
  - ICDAR Robust Reading Challenge for more difficult English Scene Text
  - Bartosgye's Handwritten character dataset: Handwritten characters for both English and Hungarian special characters

Training Images	Validation Images	Test Images
98, 717	32,919	32,915

Model Pipeline

- A CNN was used with two Fully connected layers at the end

- Data augmentation was done on the training data using:
  - Random Greyscaling of image
  - Random rotation of +22.5 to -22.5 degrees

Data Preparation
- Train
- Validation
- Test

Data Augmentation
- Grey-scaling
- Rotation
- Only on train data

Data Loaders
- Train Loader
- Val Loader
- Test Loader

Model
- Embedding Layer
- Convolution Blocks
  - Average Pool
  - Convolution 2D
  - RELU Activation
  - Batch Normalization
- FCNN 1
- FCNN 2
- Softmax

Validation
- Hyper Parameter Tuning

Testing

Fine Tuning

Model Hyperparameters to consider

- **Embedding depth:** Initial Embedding depth gets doubled and the dimensions of the input channels get halved at every layer, so the embedding depth determines model size

- **Image size:** Since multiple sources of images are used, image size becomes vital

- **Batch Norm or not:** Should batch norm be used or not?

- **Learning Rate**

- **Batch size**