THE OPPORTUNITIES PRESENTED BY MACHINE LEARNING APPLIED IN EDUCATION AND THE ENTAILING CHALLENGES

Oskar Thaeter

Referenzfach: Informatik

Bezugsfach: Englisch

Prüfer: Jan Reher

STRUCTURE

Hypothesis

Machine Learning

Application

Opportunities

Challenges

Conclusion

HYPOTHESIS

AN AI REVOLUTION

- machine learning promises to revolutionize industries
- artificial intelligence is ever present in day to day life
- computing power fuels machine learning popularity
- AI relieves humans of repetitive tasks

DATA AS A RESOURCE

- 700,000 pupils take centric exams in UK
- each exam has to be marked by two individuals
- all this data is wasted
- should be used to train AI
- AI could relieve teachers of marking

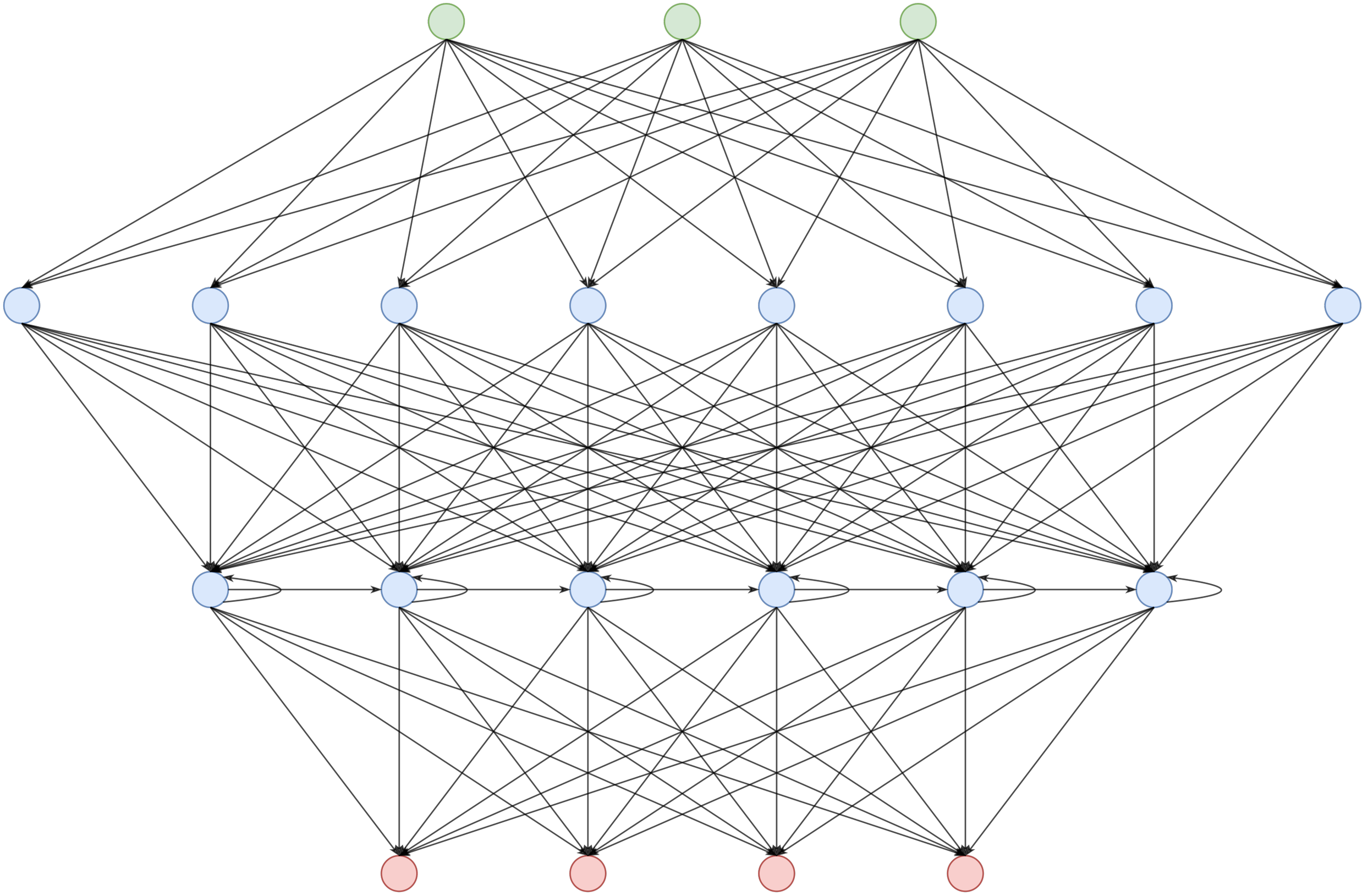
MACHINE LEARNING

WHAT IS MACHINE LEARNING ?

- attempts to recreate human intelligence → AI
- computer learns from data to carry out a task
- concept: learn from past mistakes to maximize success
- learning: process of training a mathematical model

ARTIFICIAL NEURAL NETWORKS

- network of simple information processing units (neurone)
- inspired by the neural network of a human brain
  - neurone
    - holds a value, typically between 0 and 1
    - applies an activation function to weighted input
    - feeds output via weighted connection to next neurone



APPLICATION

THE PROBLEM

- translate handwriting into machine readable data
- formally called optical character recognition(OCR)
- problem statement:
  - translate a picture of fixed size of a written word into according digital output.

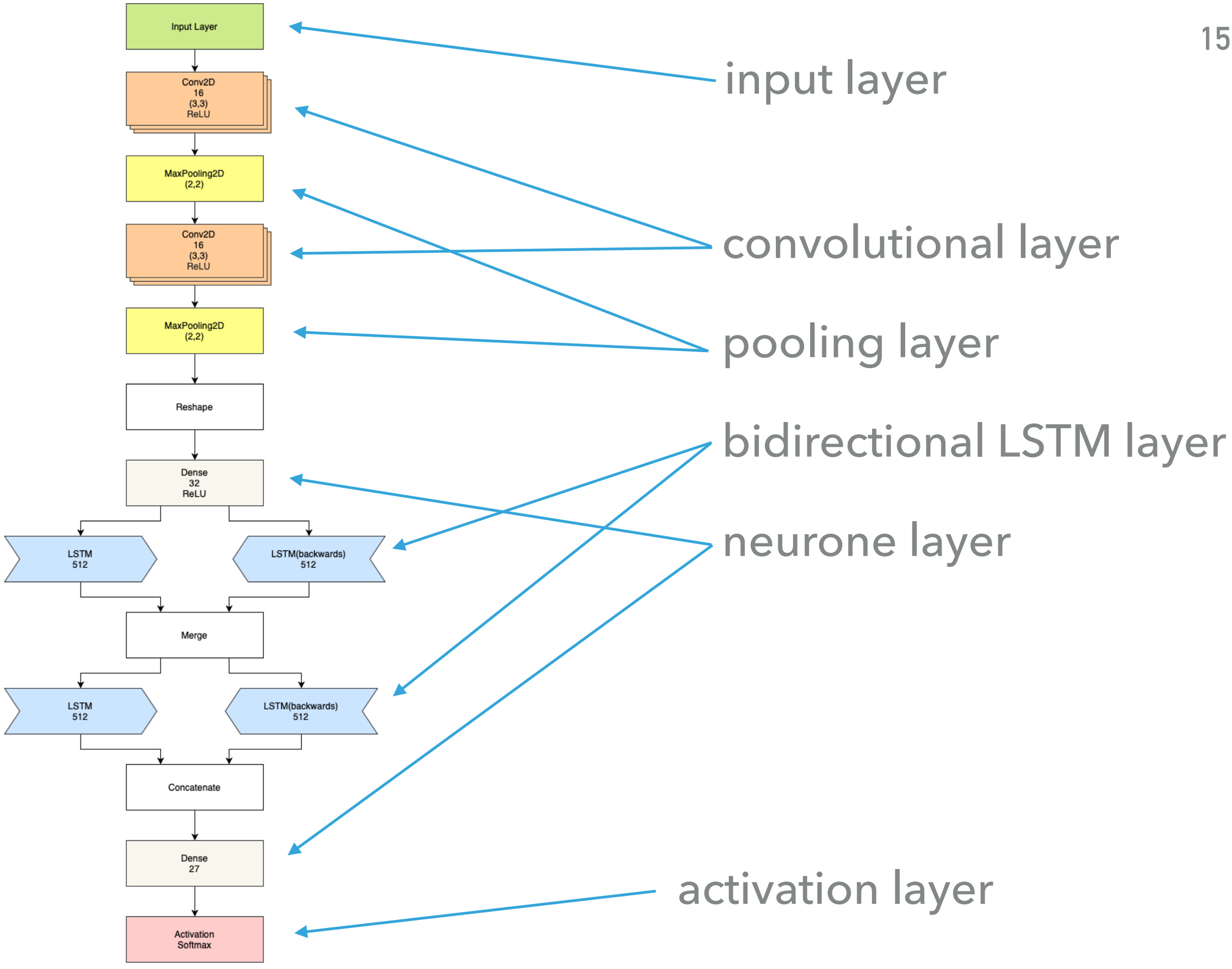
cure

DATASET

- consists of images and labels
- synthetic dataset -> almost infinite size, many variations
- no costs to produce
- more control over the data
- correctly labeled data
- 23 different fonts

MODEL

- convolutional neural network (CNN)
  - to detect features in the image
- recurrent neural network (RNN)
  - outputs a sequence of variable length
- connectionist temporal classification (CTC)
  - decodes the output into words


Input Layer
Conv2D
16
(3,3)
ReLU
MaxPooling2D
(2,2)
Conv2D
16
(3,3)
ReLU
MaxPooling2D
(2,2)
Reshape
Dense
32
ReLU
LSTM
512
LSTM(backwards)
512
Merge
LSTM
512
LSTM(backwards)
512
Concatenate
Dense
27
Activation
Softmax
input layer
convolutional layer
pooling layer
bidirectional LSTM layer
neurone layer
activation layer

BUILDING THE MODEL

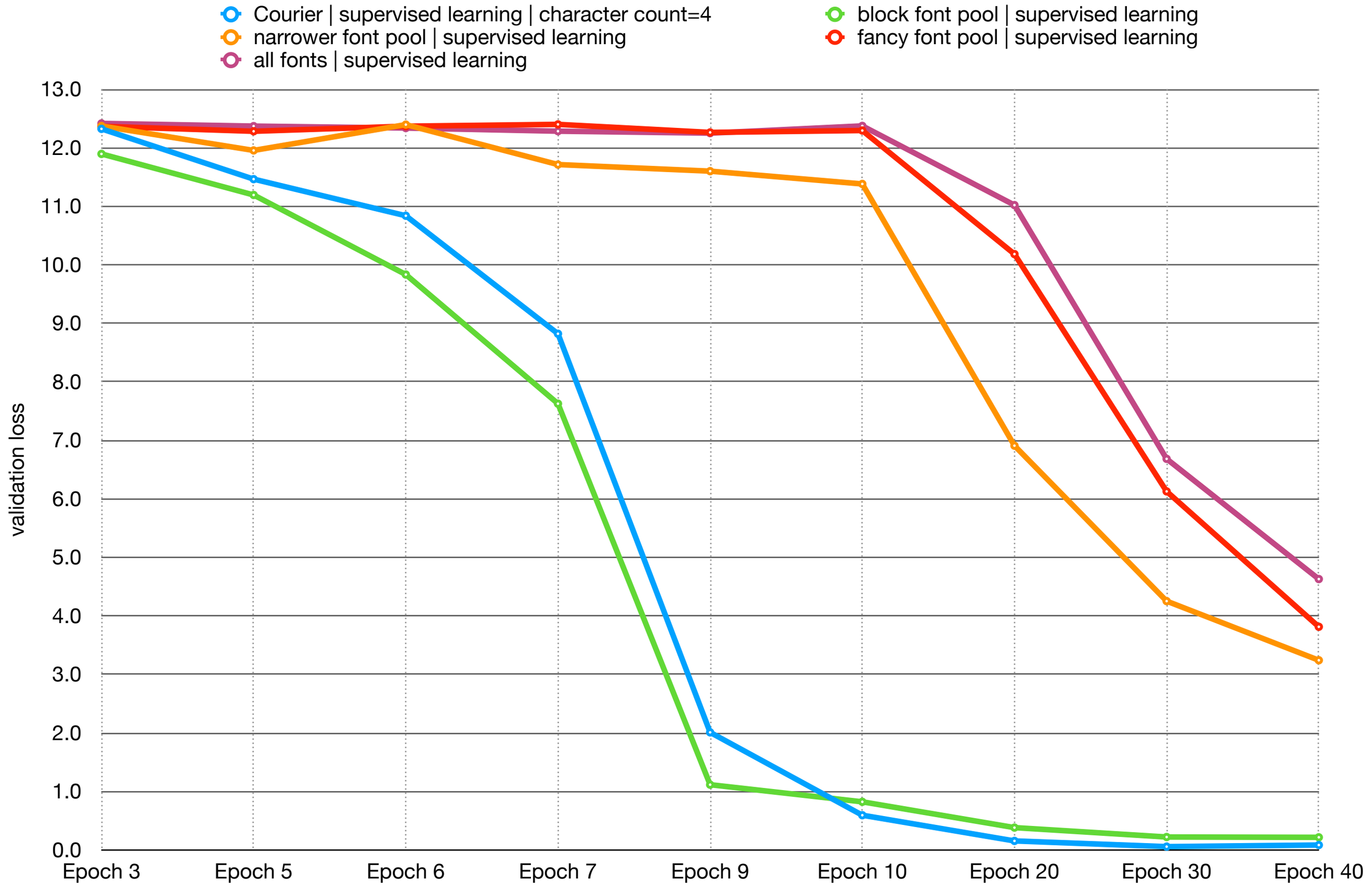
- using Python and the Keras API
- NVIDIA Tesla K80 GPU via Google Cloud Computing
  - epoch takes up to 260s depending on the model
- Cairo library to generate images
  - creates 16 000 new images every epoch

TRAINING THE MODEL

- gradient descent optimiser
  - changes weights between neurones to maximize success
- loss function: CTC loss function
  - indicates how bad a prediction is
- goal is to minimize loss

ANALYSING AND IMPROVING THE MODEL

- empirical approach
- How complex is the task?

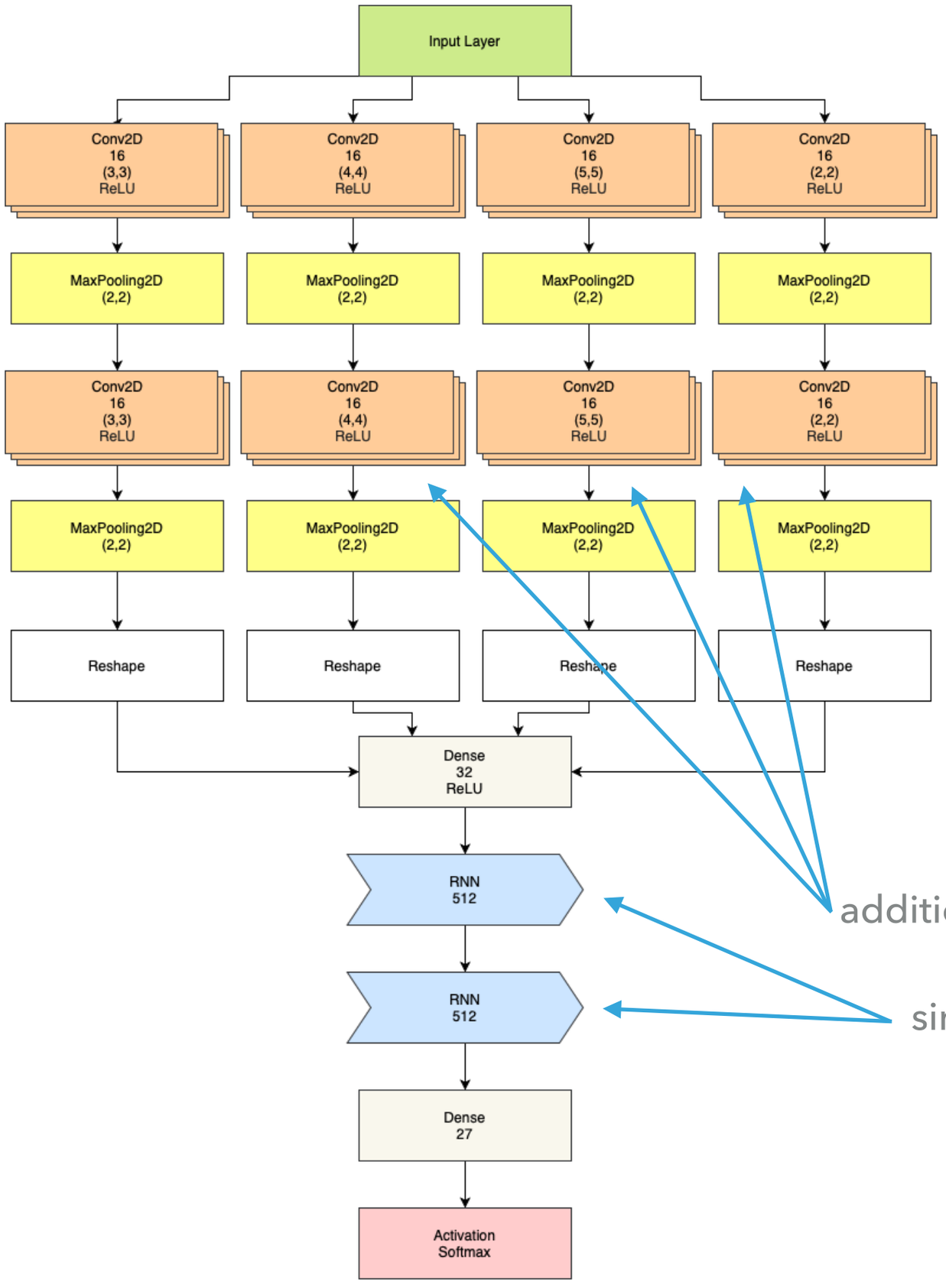


validation loss - lower is better

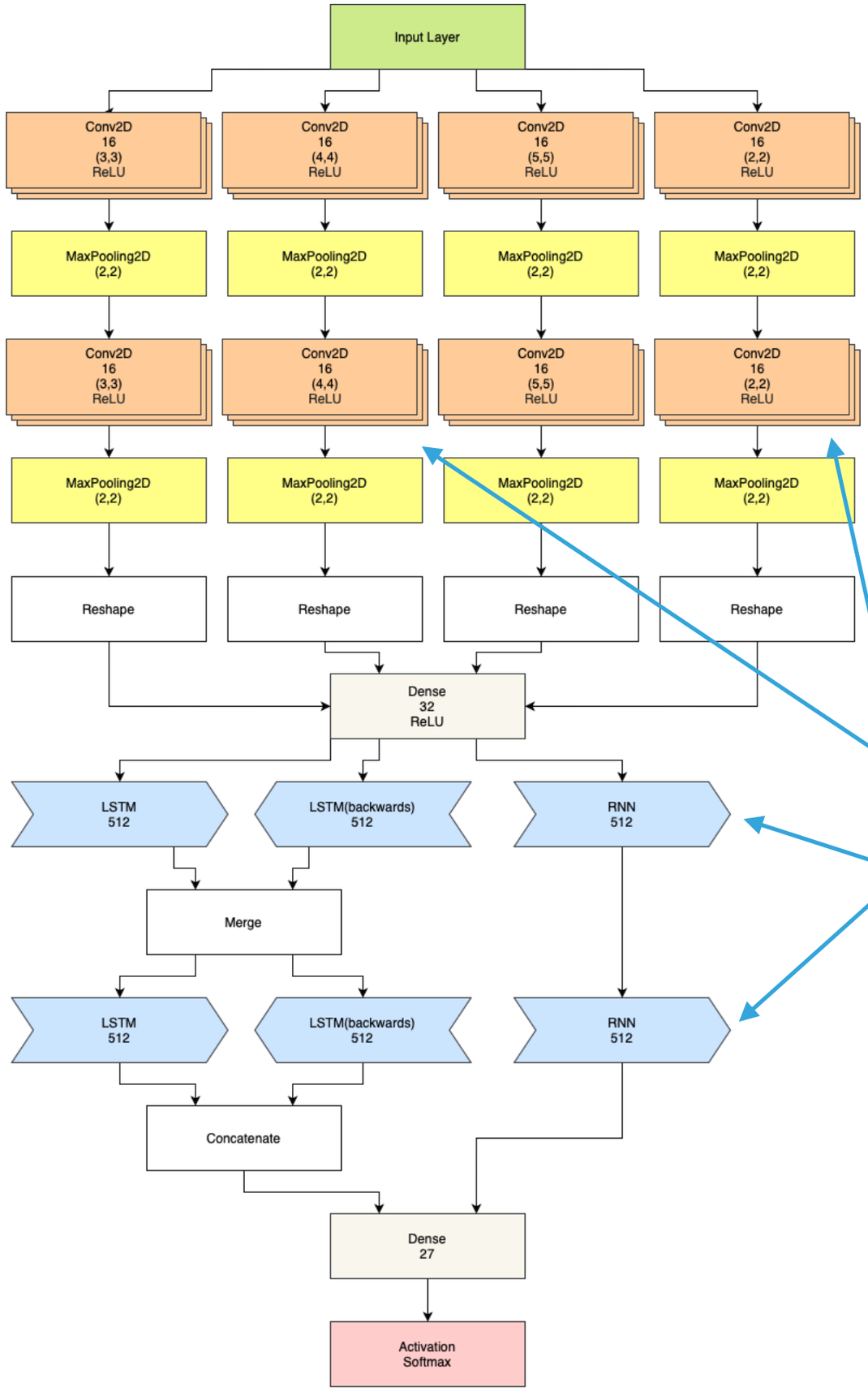
ANALYSING AND IMPROVING THE MODEL

- empirical approach
- How complex is the task?
- alternative architectures

CNN_parallel_RNN



CNN_parallel_LSTM_RNN

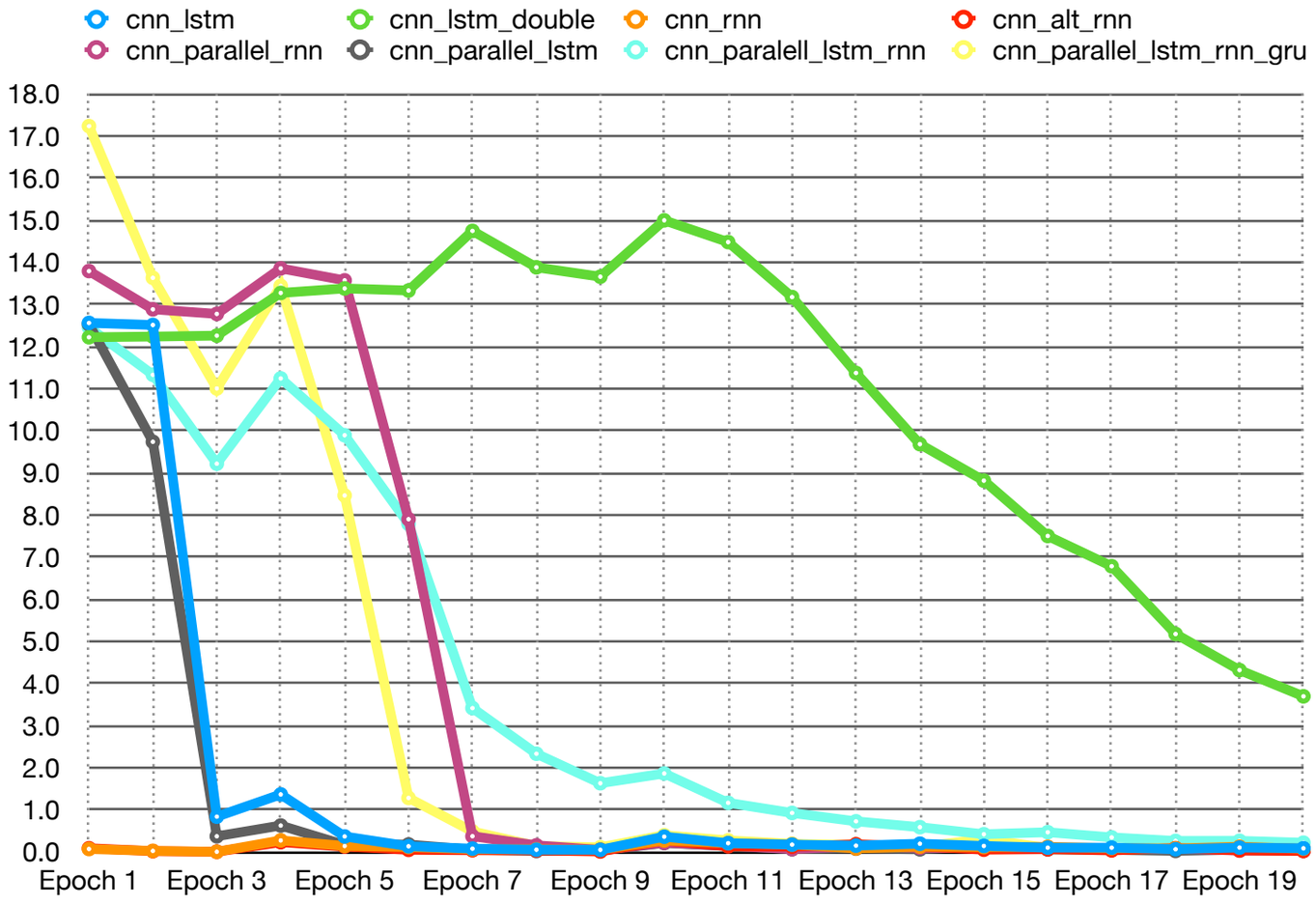


parallel convolutional stacks

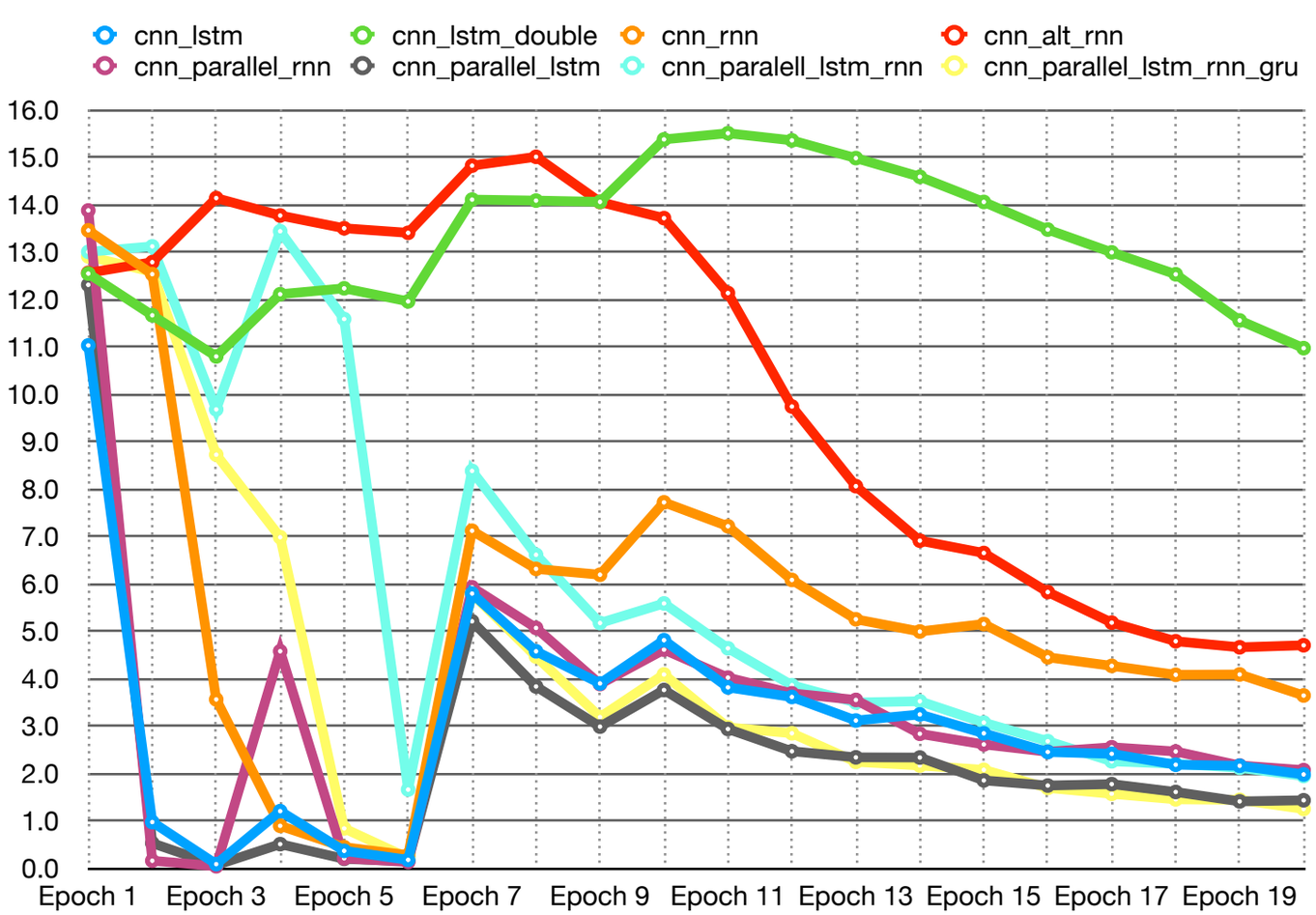
simple recurrent layers

bi LSTM layers

Courier dataset



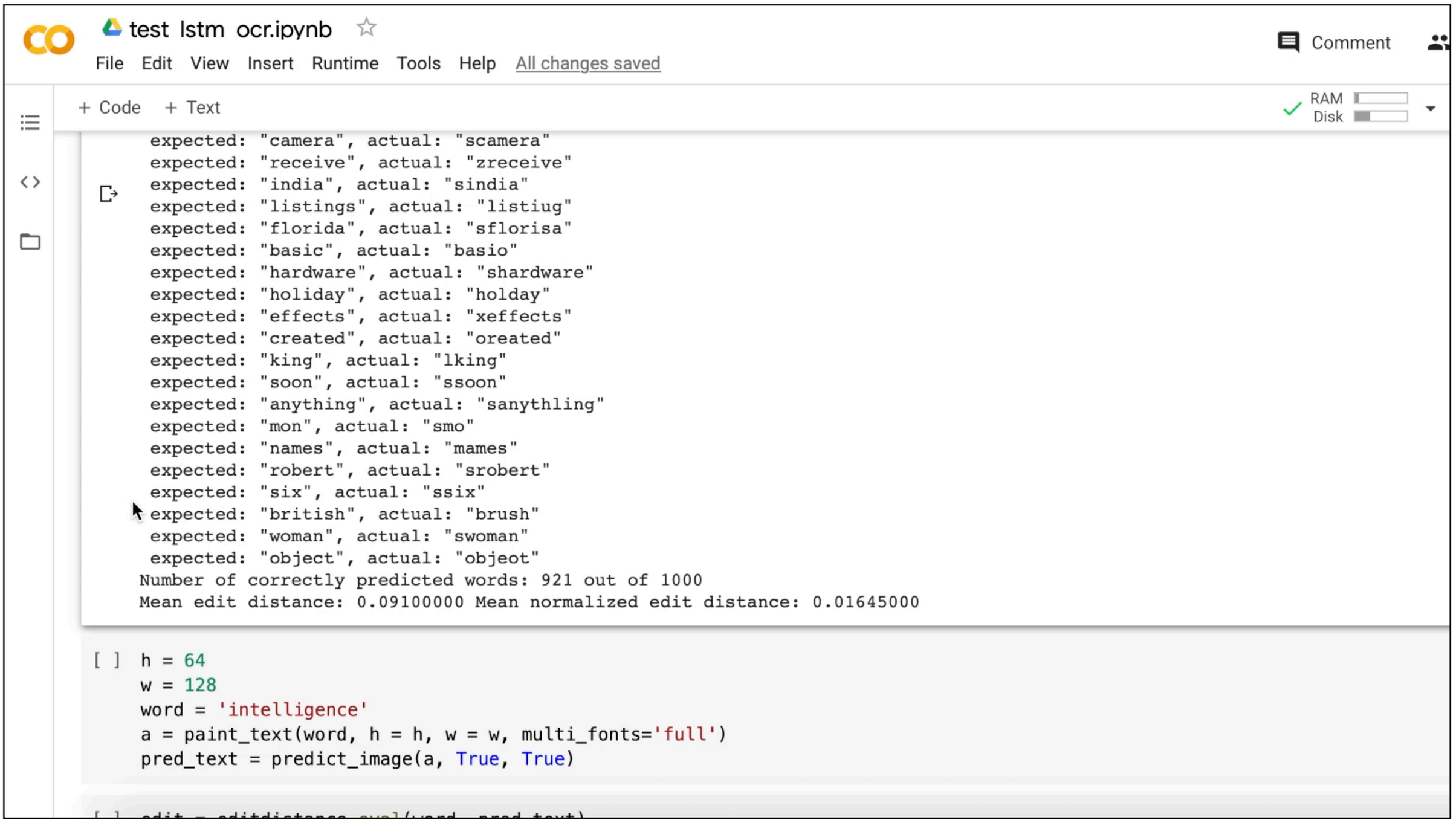
full dataset



THE FINAL MODEL

- cnn_parallel_lstm_rnn
- 60 epochs
  - validation loss: 0.5284
  - recognizes 906 out of 1000 words correctly
- 100 epochs
  - validation loss: 0.4121
  - recognizes 811 out of 1000 words correctly
- 200 epochs
  - validation loss: 0.2537
  - recognizes 923 out of 1000 words correctly

THE FINAL MODEL



THE FINAL MODEL

test lstm ocr.ipynb ☆

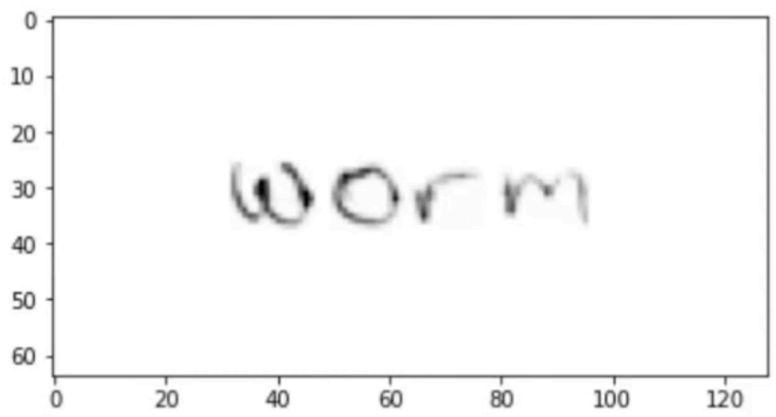
File Edit View Insert Runtime Tools Help All changes saved

Comment

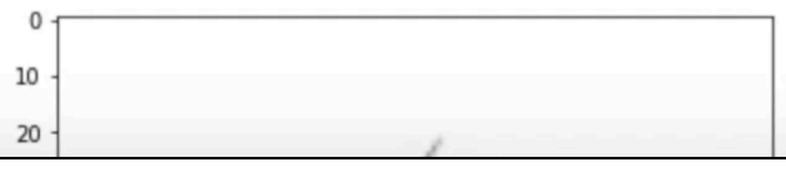
+ Code + Text

RAM Disk

```
Average time to predict word: 0.8133319854736328 seconds
False predictions:
 expected: "worm", actual: "vvovn"
 expected: "employee", actual: "sxployee"
 expected: "work", actual: "vvok"
 expected: "danger", actual: "clungcr"
 expected: "favour", actual: "yaveso"
 expected: "enigma", actual: "zryyya"
 expected: "world", actual: "vorlds"
 expected: "would", actual: "bsjbnlss"
 expected: "cure", actual: "zttsst"
Number of correctly predicted words: 1 out of 10
Mean edit distance: 3.70000000 Mean normalized edit distance: 0.74666667
```



"vvovn"



OPPORTUNITIES

DATA DEMOCRATISATION

- advances in ML should be open source
- sharing data can help decision makers
- OCR can work across borders
- datasets are the most expensive part

CURRENT TECHNOLOGY

- Automated Students Assessment Prize
- automated essay scoring(AES)
  - feature-engineered models
  - end-to-end models

CHALLENGES

TECHNOCHAUVINISM

"UNWAVERING FAITH THAT IF THE WORLD JUST USED MORE COMPUTERS, AND USED THEM PROPERLY, SOCIAL PROBLEMS WOULD DISAPPEAR AND WE'D CREATE A DIGITALLY ENABLED UTOPIA."

Meredith Broussard

TECHNOCHAUVINISM

- technology can not solve everything
- social issues persist
  - unequal access to digital infrastructure
  - low education budgets

THE NEED FOR HUMAN EXPERIENCE

- “It is important to be open minded and positive when marking scripts.”
- demands consciousness
- marking english papers too complex for current narrow AI

BIAS

- “state-of-the-art referred to […] is, largely, simply counting words”
- current AES favours sophisticated vocabulary and text length over content and coherence

CONCLUSION

DOES MACHINE LEARNING OFFER A VIABLE SOLUTION TO EDUCATIONAL ISSUES?

- use of AI is inevitable
- legal, cultural and technical issues
- marking is not the biggest problem of education today
- if marking is a problem, demands on teachers are too high
- ML can fight symptoms, can not tackle root causes

LITERATURE

Moore, Gordon E. (1965): “Cramming more components onto integrated circuits”
(Electronics, Volume 38, Number 8, April 19, 1965)
URL: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf
[Retrieved 09/12/2019]

Kelleher. John D.: Deep Learning
Cambridge, Massachusetts: The MIT Press, 2019

Chung et al. (2014): “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”
page 7, scientific publication
URL: https://arxiv.org/pdf/1412.3555v1.pdf
[Retrieved 01/02/2020]
preprint scientific paper comparing different types of recurrent units in RNN

Hochreiter, Sepp und Schmidhuber, Jürgen: “Long Short-Term Memory”
(Neural Computation, 9(8):1735–1780, 1997)
scientific publication
URL: https://www.bioinf.jku.at/publications/older/2604.pdf
[Retrieved 06/02/2020]
scientific paper introducing Long Short-Term Memory

Cho et al. (2014): “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”
scientific publication
URL: https://arxiv.org/abs/1406.1078
[Retrieved 01/02/2020]

Radford et al. (2019): “Language Models are Unsupervised Multitask Learners”
URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
[Retrieved 03/01/2020]
scientific publication

Jiawei Liu, Yang Xu, und Yaguang Zhu (2019): “Automated Essay Scoring based on Two-Stage Learning”
page 1, scientific publication
URL: https://arxiv.org/pdf/1901.07744.pdf
[Retrieved 04/02/2020]
preprint scientific paper proposing a Two-Stage Learning Framework, using feature-engineered and end-to-end AES methods

Fei Dong and Yue Zhang: “Automatic Features for Essay Scoring – An Empirical Study”
(Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1072–1077, Austin, Texas, November 1-5, 2016.© 2016 Association for Computational Linguistics; 2016)
URL: https://www.aclweb.org/anthology/D16-1115.pdf
[Retrieved 04/02/2020]
scientific paper on automatic feature learning in automatic essay scoring

Broussard, Meredith: Artificial Unintelligence: How Computer Misunderstand the World
Cambridge, Massachusetts: The MIT Press, 2018

Les Perelman: “When ‘the state of the art’ is counting words”
(Assessing Writing 21 (2014) 104–111; 2014)
URL: http://lesperelman.com/wp-content/uploads/2015/09/Perelman-State-of-the-Art-is-Counting-Words.pdf
[Retrieved 04/02/2020]
research article on the current state of automated essay scoring

SOURCES

Klovert, Heike; Laurenz, Nike und Quecke, Franca (2019): “Warum in Deutschland so viele Lehrer fehlen: Grund- und Berufsschulen spüren es besonders stark: Bundesweit fehlen Lehrer. Woran liegt das? Wie steuern die Länder gegen? Die wichtigsten Antworten im Überblick.”
URL: https://www.spiegel.de/lebenundlernen/schule/lehrer-mangel-in-deutschland-warum-so-viele-paedagogen-fehlen-und-wo-a-1280209.html
[Retrieved 09/12/2019]
article in “Spiegel Online”

Zorn, Dr. Dirk (2019): “Lehrermangel in Grundschulen bis 2030 größer als bislang erwartet”
URL: https://www.bertelsmann-stiftung.de/de/themen/aktuelle-meldungen/2019/september/lehrermangel-in-grundschulen-bis-2030-groesser-als-bislang-erwartet/
[Retrieved 09/12/2019]
“Bertelsmann Stiftung” study
“Die Bewältigung des Lehrermangels ist eine Herkulesaufgabe. Die Zahl der Schülerinnen und Schüler wächst dynamischer als angenommen.”

Statistisches Bundesamt (2018): “Absolventen/Abgänger nach Abschlussart und Geschlecht”
URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/liste-absolventen-abgaenger-abschlussart.html;jsessionid=36A47054D1EEB35C7D3C66BC068CF4D0.internet711
[Retrieved 03/01/2020]

BBC News (22.08.2019): “GCSEs 2019: Where are we at with new grades 9-1?”
URL: https://www.bbc.com/news/education-48993830
[Retrieved 03/01/2020]

The Office of Qualifications and Examinations Regulation (2018): “Entries for GCSE, AS and A level: Summer 2018 exam series”
URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/712450/Report_-_summer_2018_exam_entries_GCSEs_Level_1_2_AS_and_A_levels.pdf
[Retrieved 03/01/2020]
annually published, official statistics

AQA awarding body: “How Exams Work”
URL: https://www.aqa.org.uk/about-us/what-we-do/getting-the-right-result/how-exams-work
[Retrieved 03/01/2020]

Cambridge Dictionary: “machine learning”
URL: https://dictionary.cambridge.org/dictionary/english/machine-learning
[Retrieved 08/12/2020]

Keras Documentation: “About Keras layers”
URL: https://keras.io/layers/about-keras-layers/
[Retrieved 08/12/2020]

Keras Documentation: “Keras: The Python Deep Learning library”
URL: https://keras.io/#you-have-just-found-keras
[Retrieved 01/02/2020]

Keras Documentation: “Optical character recognition”
URL: https://keras.io/examples/image_ocr/
[Retrieved 01/02/2020]
exemplary python code using Keras

SOURCES

Keras Documentation: “Guide to the Functional API”
URL: https://keras.io/getting-started/functional-api-guide/
[Retrieved 01/02/2020]

Google Machine Learning Crash Course: “Descending into ML: Training and Loss”
URL: https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss
[Retrieved 05/02/2020]
part of an online course on machine learning by Google

OpenAI (2018): “OpenAI Charter”
URL: https://openai.com/charter/
[Retrieved 03/01/2020]

Marr, Bernard (2017): “What Is Data Democratization? A Super Simple Explanation And The Key Pros And Cons”
(Forbes; 24.07.2017)
URL: https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/#716c97fb6013
[Retrieved 03/01/2020]

The Hewlett Foundation (2012): “The Hewlett Foundation: Automated Essay Scoring”
URL: https://www.kaggle.com/c/asap-aes/data
[Retrieved 03/01/2020]
datasource for ASAP AES competition, this dataset is most often used in automated essay scoring systems

A-Level English Language 7702/1 Language, The Individual and Society (New) Final Mark Scheme June 2017
page 3
“It is important to be open minded and positive when marking scripts.”
URL: https://filestore.aqa.org.uk/sample-papers-and-mark-schemes/2017/june/AQA-77021-W-MS-JUN17.PDF
[Retrieved 04/02/2020]

The Learning Network: “What Students Are Saying About How to Improve American Education: An international exam shows that American 15-year-olds are stagnant in reading and math. Teenagers told us what’s working and what’s not in the American education system.”
(The New York Times; 19.12.2019)
URL: https://www.nytimes.com/2019/12/19/learning/what-students-are-saying-about-how-to-improve-american-education.html
[Retrieved 04/02/2020]
The New York Times article about the results of a prompt asking students how American education could be improved in light of stagnant performance of American teenagers in reading and math according to a study by the Program for International Student Assessment

CODE SOURCES

Henry, Mike (2019)
URL: https://github.com/keras-team/keras/blob/master/examples/image_ocr.py
[Retrieved 06/02/2020]
basis of the training model code, “Image OCR” example showing off Keras API

Zhang, Chengwei (2018)
URL: https://github.com/Tony607/keras-image-ocr/blob/master/image-ocr.ipynb
[Retrieved 06/02/2020]
partly used for testing code, Jupyter notebook on GitHub

Scheidl, Harald (2019)
URL: https://github.com/githubharald/CTCDecoder/blob/master/src/BestPath.py
[Retrieved 06/02/2020]
CTC decoding algorithm (Best Path) implementation

Scheidl, Harald (2019)
URL: https://github.com/githubharald/CTCDecoder/blob/master/src/BeamSearch.py
[Retrieved 06/02/2020]
complex CTC decoding Algorithmus (Beam Search), which promises better performance than Best Path

U. Marti und H. Bunke. (2002)
URL: http://www.fki.inf.unibe.ch/databases/iam-handwriting-database/iam-handwriting-database#icdar02
[Retrieved 06/02/2020]
The IAM-database: An English Sentence Database for Off-line Handwriting Recognition.
Int. Journal on Document Analysis and Recognition, Volume 5, pages 39 - 46, 2002
dataset consisting off 115 320 images, used in final test of model