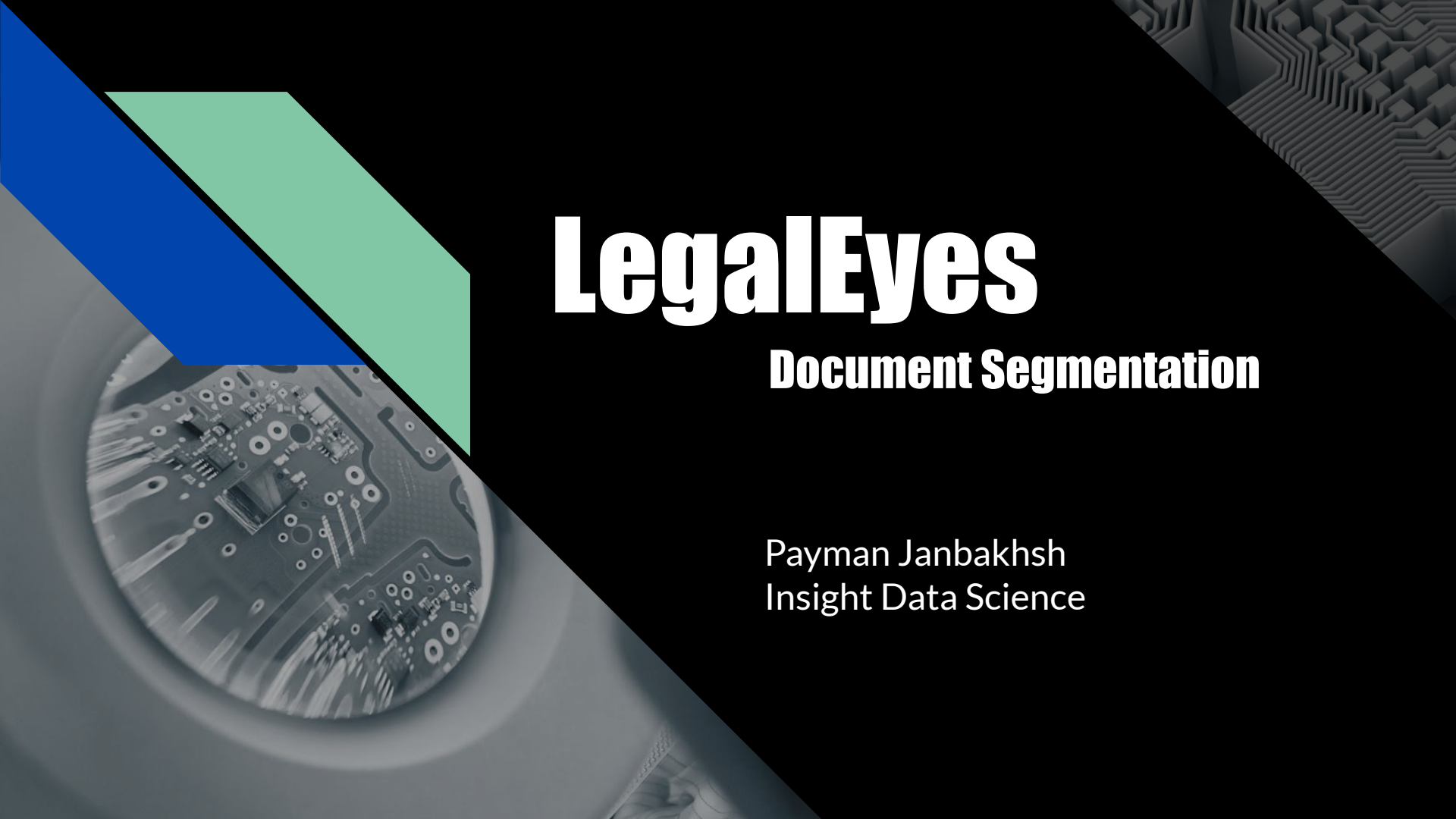


LegalEyes

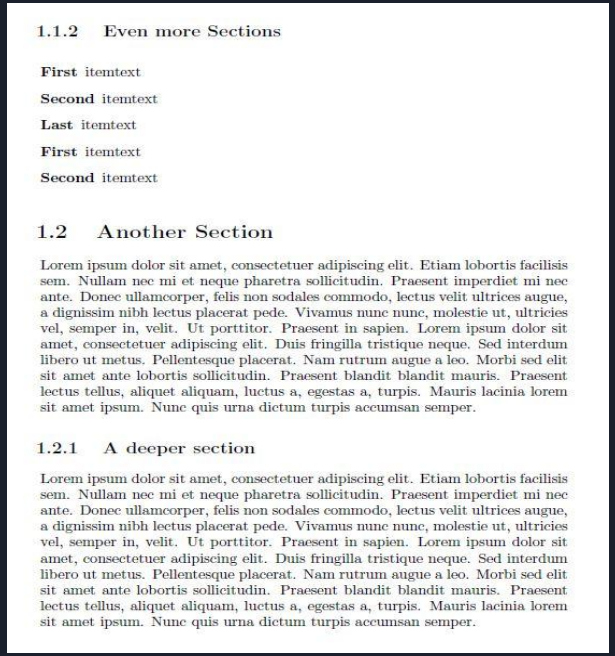
Document Segmentation

Payman Janbakhsh
Insight Data Science

Consulting Project

Automate information extraction from legal documents

HeaderFooter
Title
Body
List
Tables
Graphs/Charts



1.1.2 Even more Sections

First itemtext

Second itemtext

Last itemtext

First itemtext

Second itemtext

1.2 Another Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.2.1 A deeper section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Another Application

Your client wants you to recommend whether a new project should be funded or not

Dr. Zamin Ali Khan *et al*, International Journal of Computer Science and Mobile Computing, Vol.8 Issue.11, November- 2019, pg. 81-86

Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 11, November 2019, pg.81 – 86

Novel Approach to Design and Implement a Multi-Language Convertor using Machine Learning Techniques

Dr. Zamin Ali Khan[1]; M. Haris[2]; Syed Rizwan[3]

Department of Computer Science, IQRA UNIVERSITY, Pakistan

[1]dr.m.zamin@iqra.edu.pk

***Abstract*—This paper describes the multi-language translator application which fills in as a mediator connect for two people holding a conversation using various languages. This framework can translate any language which is written or spoken by a client to any language in which the client needs. To achieve this task, our system work on three separate models: voice to text conversion model, text to voice conversion model and translated model. Our translator will help the tourist and businessmen traveling to foreign countries with alien language which is the primary correspondence obstruction between the individual situated in various nations. It will be beneficial for them to find restaurants, toilets, hotels, libraries, etc.**

***Keywords*—Translator, Machine learning, Model, Voice-to-text (vtt), Text-to voice (ttv).**

I. INTRODUCTION

The improvements in the field of Science and Technology have made this world a Global village. Modern modes of transportation and communication have bridged the gap between different countries. Modern transportation facilities advanced enough to carry people from one side to the other side of the globe in hours and the communication is fast enough to carry the sound to anywhere in the world in seconds. Different countries have different cultures and without communication, the country is over closed that may cause disaster.

As the human history have proven again and again ,the people who don't share our culture our language we consider them as our enemies, and if we are not able to understand others, and express our feeling and emotions the xenophobia take place [2].

The difference of language proves to be the biggest barrier in bringing this inter connected world further closer to each other where people are bound with each other through economics and commercial interest [3].

So it has become very important that we come up with effective ways to break communication barriers between the people speaking different languages. This application will help break the communication barriers for people travelling abroad to foreign countries. It will be helpful for students, Businessmen and people who like travelling to new destinations. This translator will help people in transportion through different destinations, visiting historical places, reading the local newspapers and signboards and understanding the true sentiment of the people. Often in foreign country, People get stuck at strange places because they are not familiar with the local language and it is impossible for them to read the different enlisted bus routes. With this translator you can either write in the sentences from Urdu or any foreign language and it will translate that sentence in to your desired language. Furthermore this translator also works on voice notes meaning you can

NLP analysis

Entity extraction?
Text classification?

...

NLP on the entire doc ?

Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 11, November 2019, pg.81 – 86

Novel Approach to Design and Implement a Multi-Language Convertor using Machine Learning Techniques

Dr. Zamin Ali Khan[1]; M. Haris[2]; Syed Rizwan[3]
Department of Computer Science, IQRA UNIVERSITY, Pakistan
[1]dr.m.zamin@iqra.edu.pk

***Abstract*—This paper describes the multi-language translator application which fills in as a mediator connect for two people holding a conversation using various languages. This framework can translate any language which is written or spoken by a client to any language in which the client needs. To achieve this task, our system work on three separate models: voice to text conversion model, text to voice conversion model and translated model. Our translator will help the tourist and businessmen traveling to foreign countries with alien language which is the primary correspondence obstruction between the individual situated in various nations. It will be beneficial for them to find restaurants, toilets, hotels, libraries, etc.**

***Keywords*—Translator, Machine learning, Model, Voice-to-text (vtt), Text-to voice (ttv).**

I. INTRODUCTION

The improvements in the field of Science and Technology have made this world a Global village. Modern modes of transportation and communication have bridged the gap between different countries. Modern transportation facilities advanced enough to carry people from one side to the other side of the globe in hours and the communication is fast enough to carry the sound to anywhere in the world in seconds. Different countries have different cultures and without communication, the country is over closed that may cause disaster.

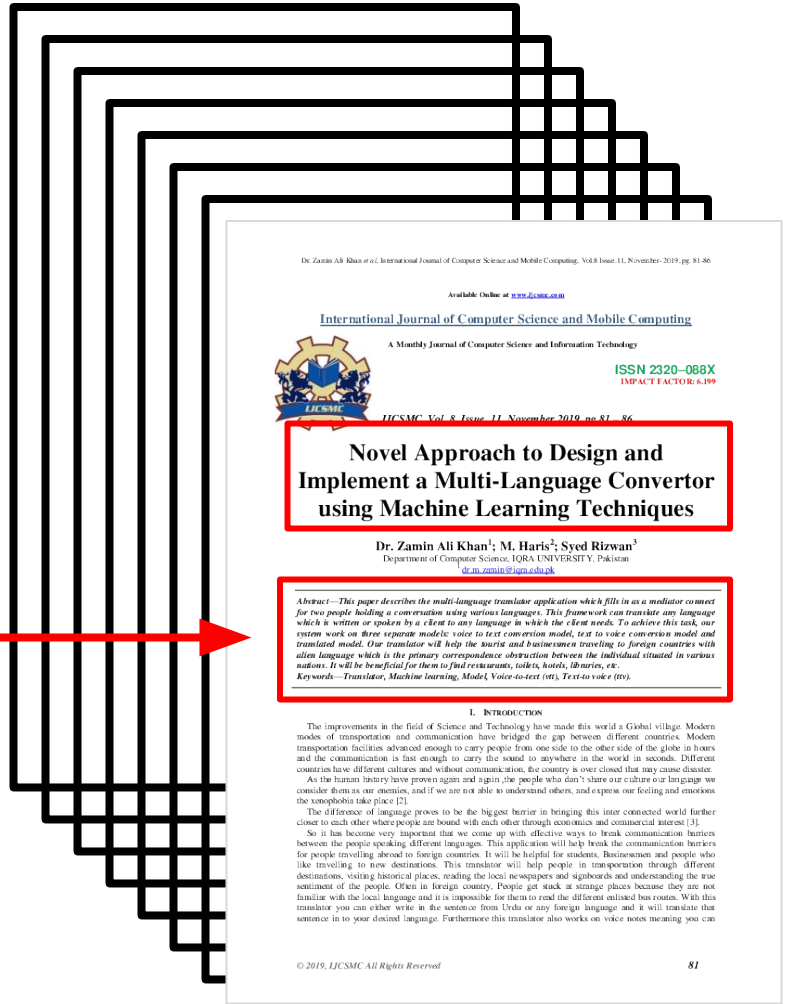
As the human history have proven again and again ,the people who don't share our culture o.r language we consider them as our enemies, and if we are not able to understand others, and express our feeling and emotions the xenophobia take place [2].

The difference of language proves to be the biggest barrier in bringing this inter connected world further closer to each other where people are bound with each other through economics and commercial interest [3].

So it has become very important that we come up with effective ways to break communication barriers between the people speaking different languages. This application will help break the communication barriers for people travelling abroad to foreign countries. It will be helpful for students, Businessmen and people who like travelling to new destinations. This translator will help people in transportion through different destinations, visiting historical places, reading the local newspapers and signboards and understanding the true sentiment of the people. Often in foreign country, People get stuck at strange places because they are not familiar with the local language and it is impossible for them to read the different enlisted bus routes. With this translator you can either write in the sentences from Urdu or any foreign language and it will translate that sentence in to your desired language. Furthermore this translator also works on voice notes meaning you can

NLP analysis

NLP to a particular segment ?



Dr. Zamin Ali Khan *et al*, International Journal of Computer Science and Mobile Computing, Vol.8 Issue.11, November- 2019, pg. 81-86

Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 11, November 2019, pg.81 – 86

Novel Approach to Design and Implement a Multi-Language Convertor using Machine Learning Techniques

Dr. Zamin Ali Khan[1]; M. Haris[2]; Syed Rizwan[3]
Department of Computer Science, IQRA UNIVERSITY, Pakistan
[1] dr.m.zamin@iqra.edu.pk

***Abstract*—This paper describes the multi-language translator application which fills in as a mediator connect for two people holding a conversation using various languages. This framework can translate any language which is written or spoken by a client to any language in which the client needs. To achieve this task, our system work on three separate models: voice to text conversion model, text to voice conversion model and translated model. Our translator will help the tourist and businessmen traveling to foreign countries with alien language which is the primary correspondence obstruction between the individual situated in various nations. It will be beneficial for them to find restaurants, toilets, hotels, libraries, etc.**

***Keywords*—Translator, Machine learning, Model, Voice-to-text (vtt), Text-to voice (ttv).**

I. Introduction

The improvements in the field of Science and Technology have made this world a Global village. Modern modes of transportation and communication have bridged the gap between different countries. Modern transportation facilities advanced enough to carry people from one side to the other side of the globe in hours and the communication is fast enough to carry the sound to anywhere in the world in seconds. Different countries have different cultures and without communication, the country is over closed that may cause disaster.

As the human history have proven again and again ,the people who don't share our culture our language we consider them as our enemies, and if we are not able to understand others, and express our feeling and emotions the xenophobia take place [2].

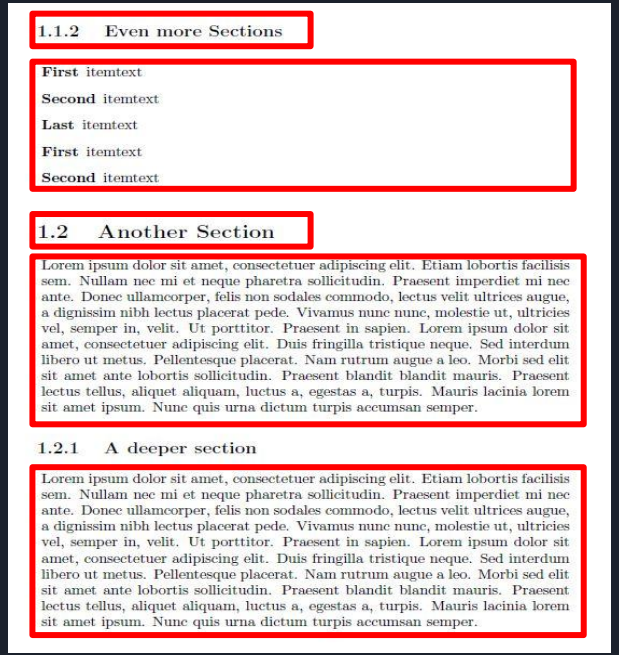
The difference of language proves to be the biggest barrier in bringing this inter connected world further closer to each other where people are bound with each other through economics and commercial interest [3].

So it has become very important that we come up with effective ways to break communication barriers between the people speaking different languages. This application will help break the communication barriers for people travelling abroad to foreign countries. It will be helpful for students, Businessmen and people who like travelling to new destinations. This translator will help people in transportion through different destinations, visiting historical places, reading the local newspapers and signboards and understanding the true sentiment of the people. Often in foreign country, People get stuck at strange places because they are not familiar with the local language and it is impossible for them to read the different enlisted bus routes. With this translator you can either write in the sentences from Urdu or any foreign language and it will translate that sentence in to your desired language. Furthermore this translator also works on voice notes meaning you can

Outside The Box

Segment & Extract

- HeaderFooter
- Title
- Body
- List
- Tables
- Graphs/Charts



1.1.2 Even more Sections

First itemtext

Second itemtext

Last itemtext

First itemtext

Second itemtext

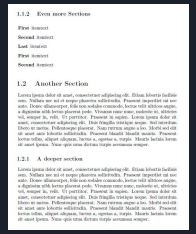
1.2 Another Section

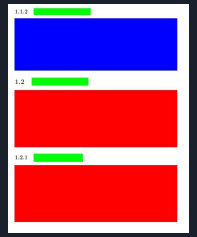
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

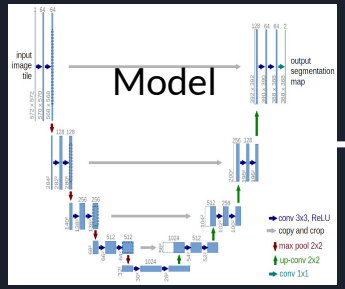
1.2.1 A deeper section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Data Cleaning







Learning Rate Adjustment

Resnet Layers Frozen/Unfrozen

Input Sizes 400, 600, 800

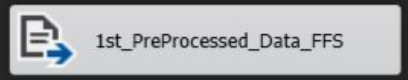
Transfer Learning

Multi_Lang OCR

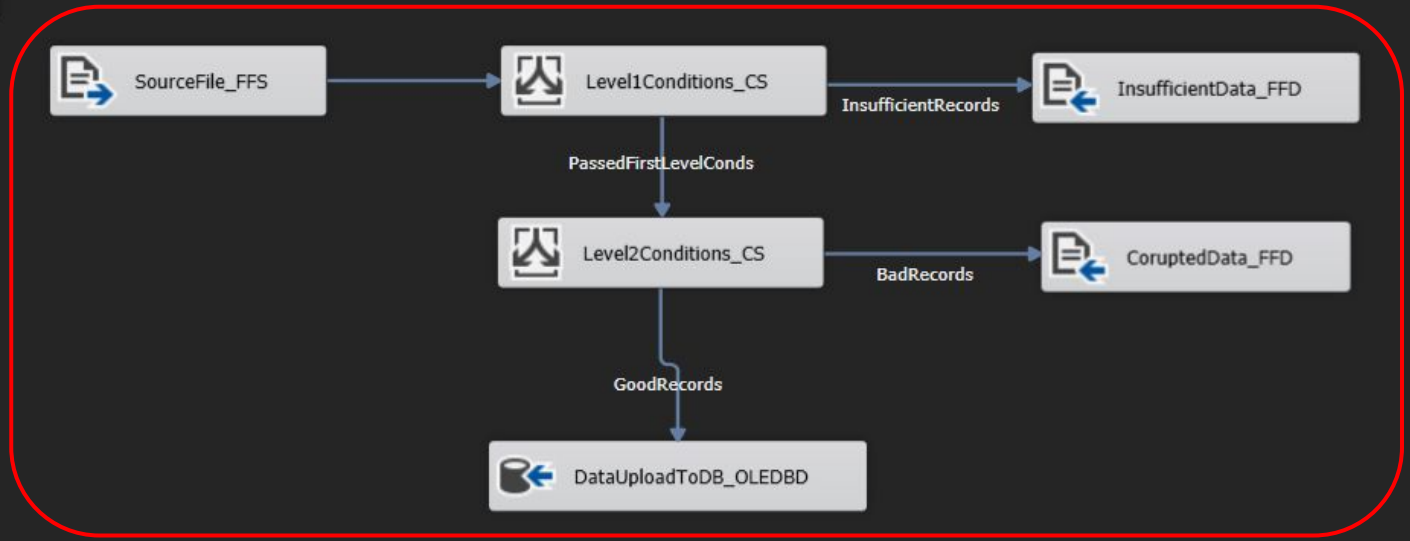
Output Docs TXT

Data Cleaning Pipeline in SSIS: TETLT

Transform (1st_preprocessing)



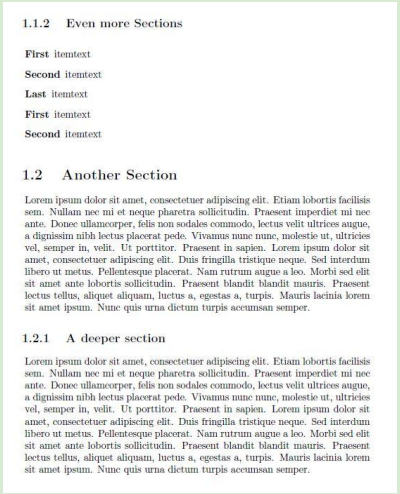
ETL



Transform (Data Types)

Data Generation/Processing

Input PDF file



1.1.2 Even more Sections

First itemtext

Second itemtext

Last itemtext

First itemtext

Second itemtext

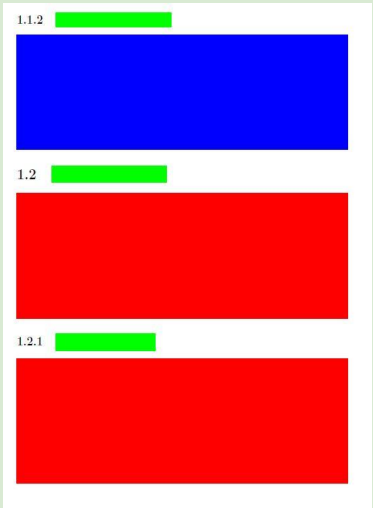
1.2 Another Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.2.1 A deeper section

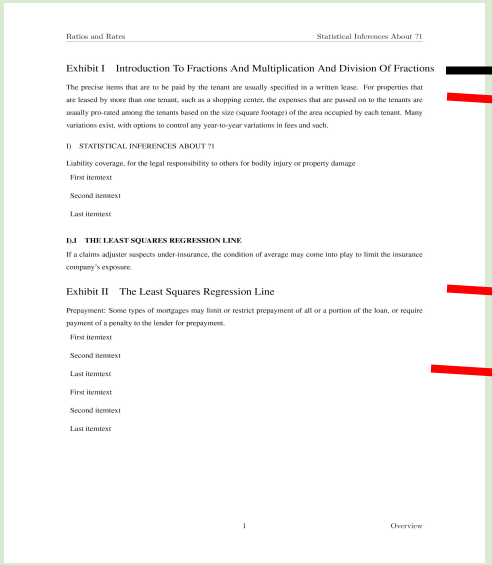
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Label PDF file



PNG images for model training

Data Cleaning



Ratios and Rates Statistical Inferences About ?1

Exhibit I Introduction To Fractions And Multiplication And Division Of Fractions

The precise items that are to be paid by the tenant are usually specified in a written lease. For properties that are leased by more than one tenant, such as a shopping center, the expenses that are passed on to the tenants are usually pro-rated among the tenants based on the size (square footage) of the area occupied by each tenant. Many variations exist, with options to control any year-to-year variations in fees and such.

I) STATISTICAL INFERENCES ABOUT ?1

Liability coverage, for the legal responsibility to others for bodily injury or property damage

- First itemtext
- Second itemtext
- Last itemtext

I).1 THE LEAST SQUARES REGRESSION LINE

If a claims adjuster suspects under-insurance, the condition of average may come into play to limit the insurance company's exposure.

Exhibit II The Least Squares Regression Line

Prepayment: Some types of mortgages may limit or restrict prepayment of all or a portion of the loan, or require payment of a penalty to the lender for prepayment.

- First itemtext
- Second itemtext
- Last itemtext
- First itemtext
- Second itemtext
- Last itemtext

1 Overview

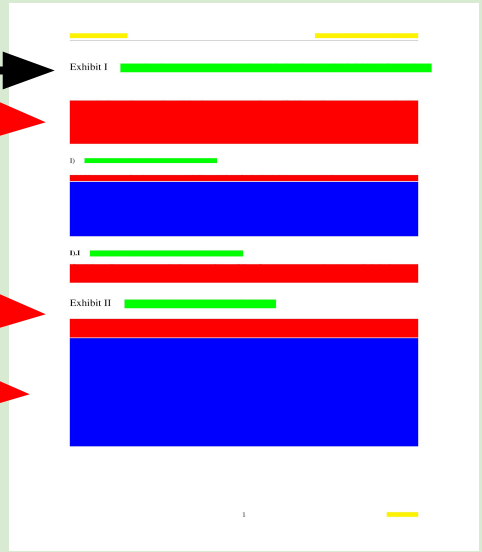


Exhibit I

I)

I).1

Exhibit II

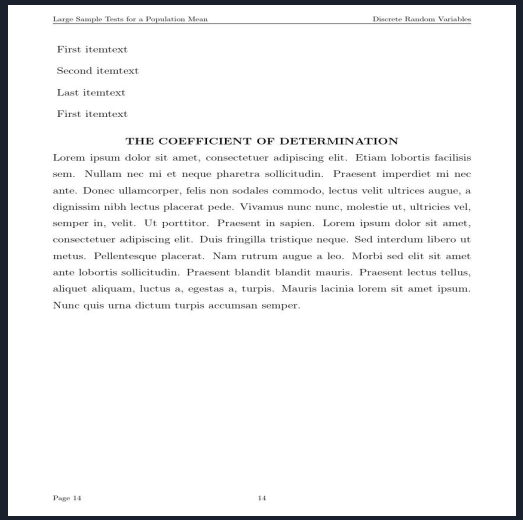
1

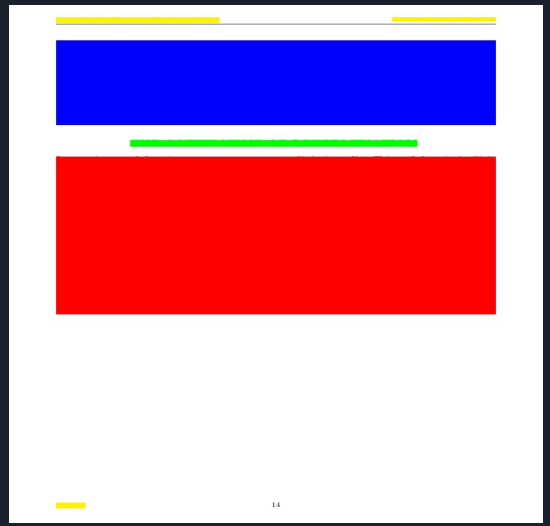
Training and Prediction

- Model was trained on 3000 pdf pages w multiple assessments of learning rate

40.00% [2/5 1:46:42<2:40:03]

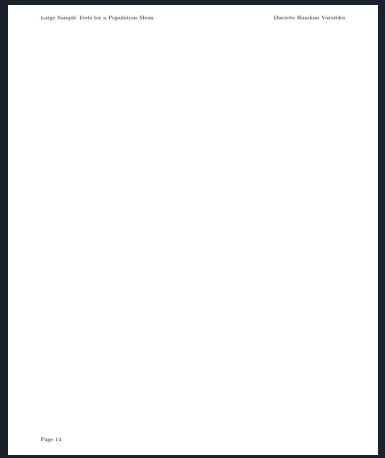
epoch	train_loss	valid_loss	Nobground_acc	Body_acc	List_acc	Header_Footer_acc	Title_acc	time
0	0.179719	0.161263	0.902312	0.923225	0.950055	0.950055	0.742894	1:04:36
1	0.111757	0.106019	0.948553	0.959866	0.990382	0.990382	0.872140	42:06



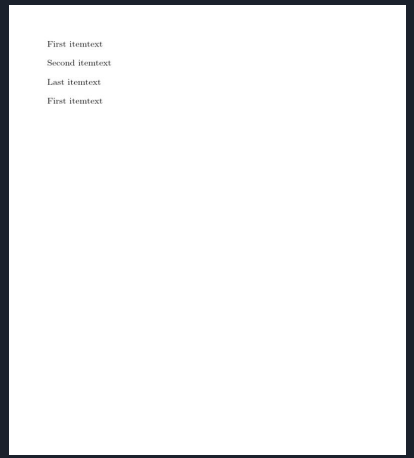


Output PNG format

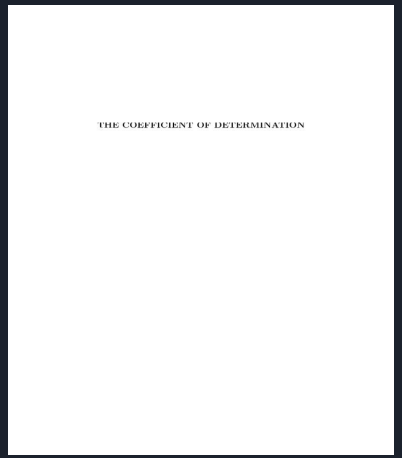
Header/Footer



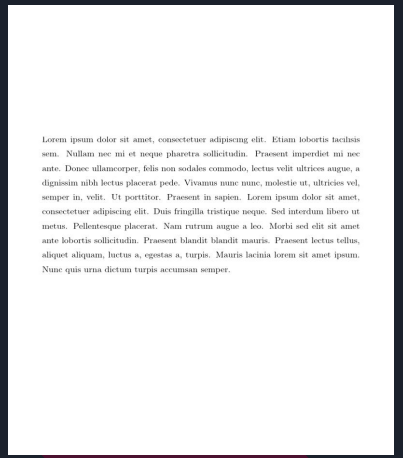
List



Title

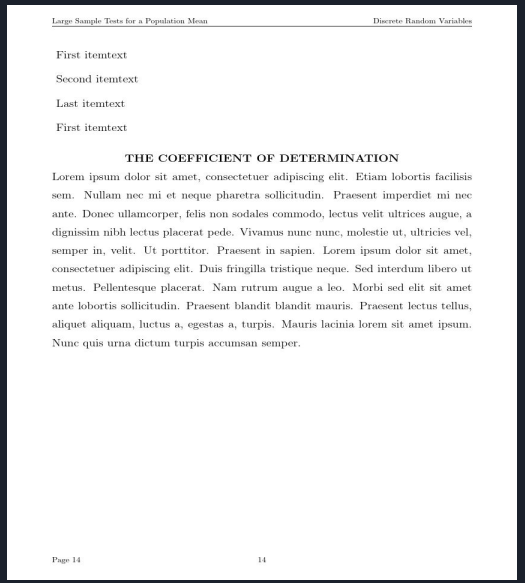


Body



OCR

PNG → TXT



First itemtext

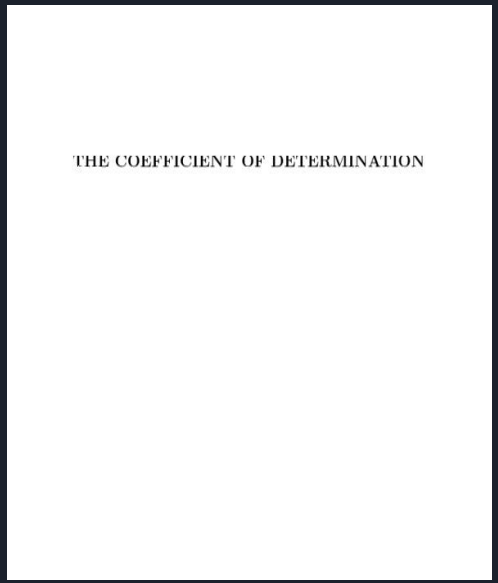
Second itemtext

Last itemtext

First itemtext

THE COEFFICIENT OF DETERMINATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



OCR

TXT

Going Forward

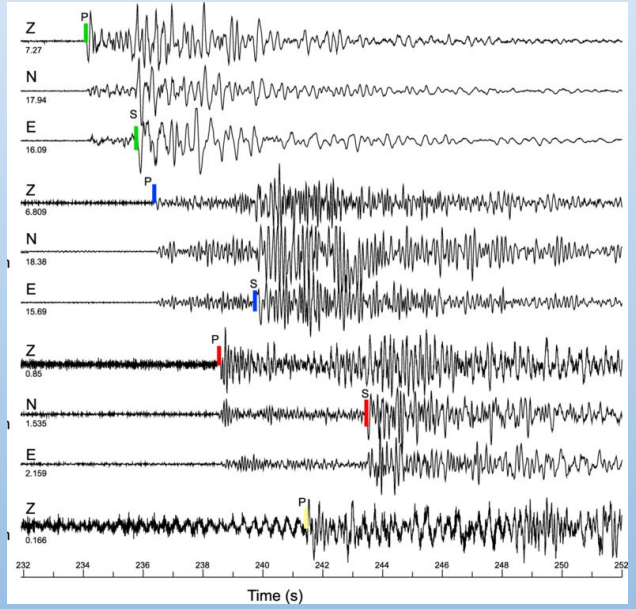
- Increasing the number of classes: Charts, Tables, Images, Figures ...
- Training the model with different doc Format/Style
- Applying other techniques to the predicted segments to increase the accuracy of the extracted text

Business Value

At large scale

- Entity-specific text or object extraction
- Removing sensitive segments from documents
- Enhancing NLP performance
- Or even text extraction from video clips for NLP analysis

Payman Janbakhsh



BSc. Applied Physics (Solid State)

MSc. Medical Physics

BSc. Geophysics (U of C)

Thesis: Estimating p-wave phase velocity using Gabor transforms

PhD. Geophysics/Deep Neural Network (U of T)

Thesis: Developed models with blocks of Autoencoders, CNN1D, LSTM

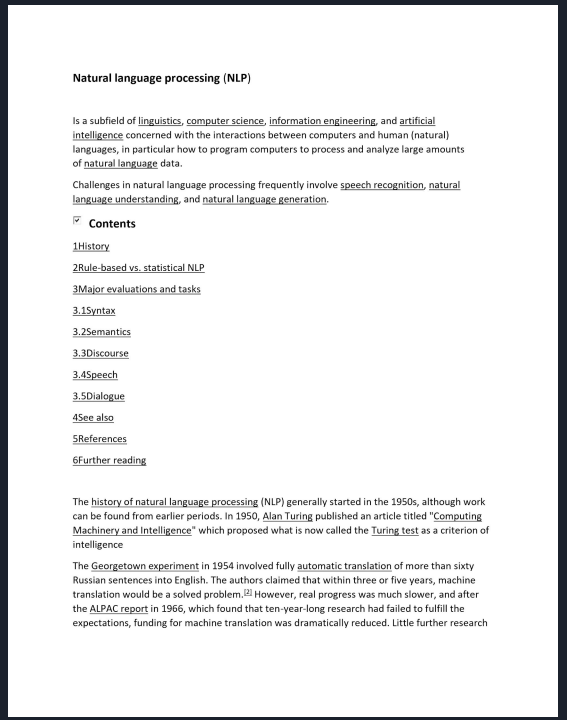
To improve Earthquake Early Detection methods.

Backup Technical slides

Testing on a different document

Model trained on 400x400 size . Must be trained on min 800x800 size

Input



Natural language processing (NLP)

Is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation.

Contents

1History

2Rule-based vs. statistical NLP

3Major evaluations and tasks

3.1Syntax

3.2Semantics

3.3Discourse

3.4Speech

3.5Dialogue

4See also

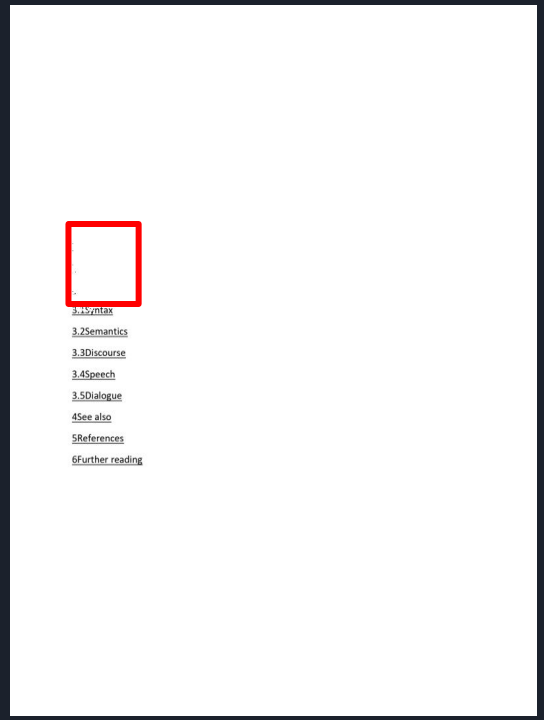
5References

6Further reading

The history of natural language processing (NLP) generally started in the 1950s, although work can be found from earlier periods. In 1950, Alan Turing published an article titled "Computing Machinery and Intelligence" which proposed what is now called the Turing test as a criterion of intelligence

The Georgetown experiment in 1954 involved fully automatic translation of more than sixty Russian sentences into English. The authors claimed that within three or five years, machine translation would be a solved problem.[2] However, real progress was much slower, and after the ALPAC report in 1966, which found that ten-year-long research had failed to fulfill the expectations, funding for machine translation was dramatically reduced. Little further research

List



3.1Syntax

3.2Semantics

3.3Discourse

3.4Speech

3.5Dialogue

4See also

5References

6Further reading

Approach

Computer Vision

1. **Image Segmentation**

2. **Using Pytorch/FastAI framework to segment and classify text objects into PNG files**

3. **Using Optical Character Recognition (OCR) to extract content from PNG to TXT files**

A. **Color code all same class objects**

B. **Now, you have finite number of objects**

C. **You have simplified the task for computer**

D. **Train your model with these labels**

E. **The output image produces class probabilities for each pixel**

Approach

Computer Vision

1. Image Segmentation

2. Using Pytorch/FastAI framework

3. Using Optical Character Recognition (OCR) to extract content from PNG to TXT files

1. Input: WxH, 3 channel images

2. Uses U-net architecture

3. Resnet34 as its encoder block

4. Class accuracies exclude bkground

5. Two transfer learning stages applied

6. Every few epochs learning rate was adjusted

Approach

Computer Vision

1. **Image Segmentation**

2. **Using Pytorch/FastAI framework**

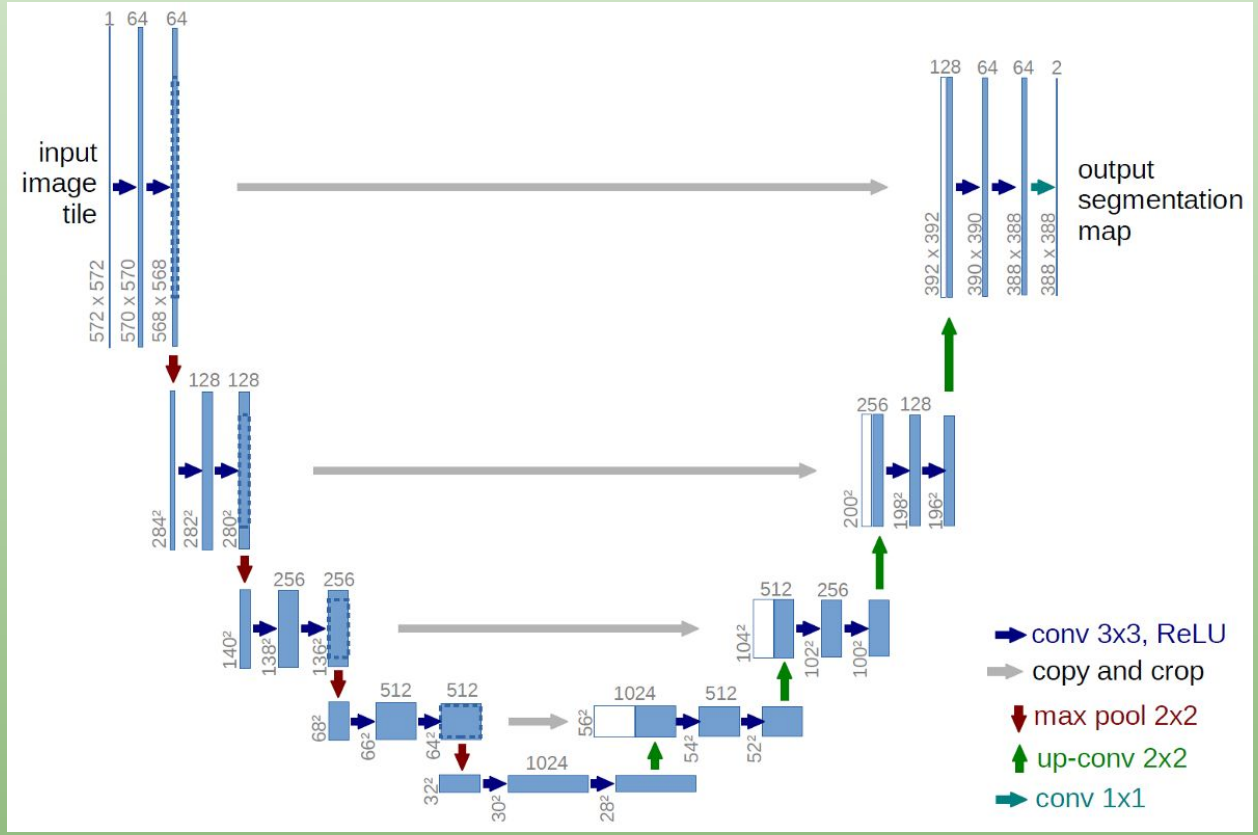
3. **Multilanguage Optical Character Recognition (OCR) to extract content from PNG to TXT files**

See challenge slide

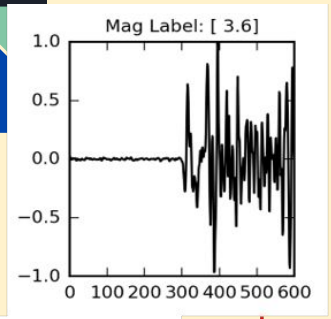
Model Architecture

- U_net architecture + Resnet34 in its encoder block

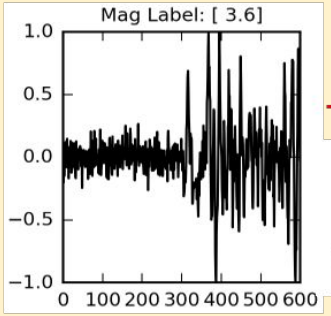
Why U_net?



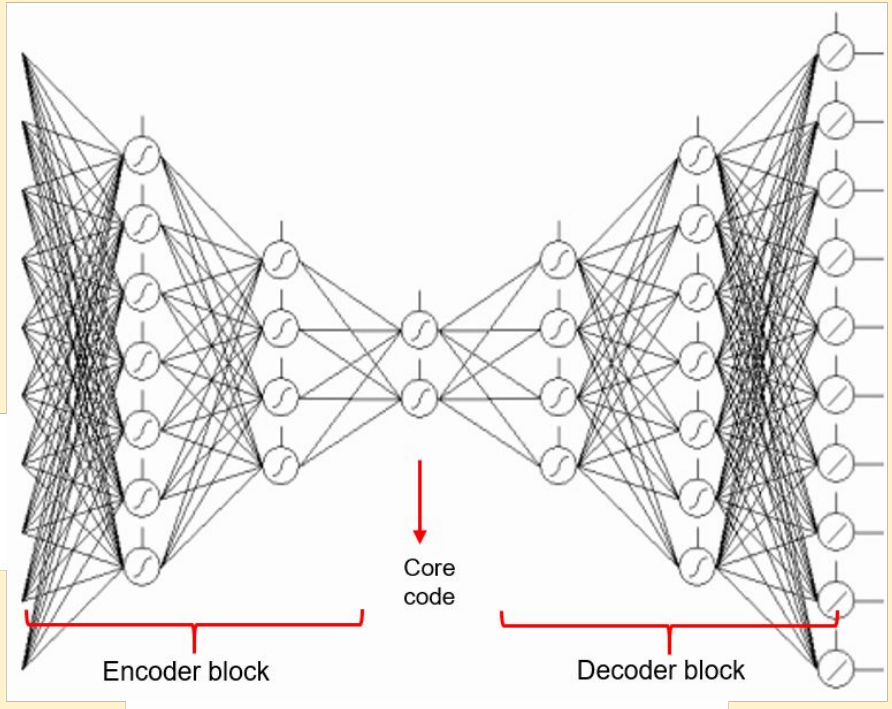
Denoise Autoencoder

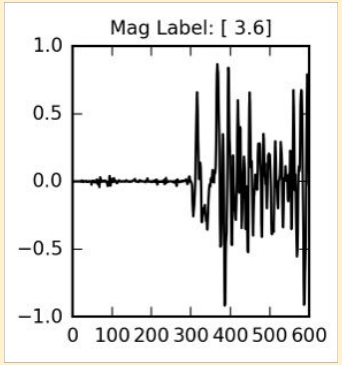


Adding noise to original signal



Noisy signal as input to autoencoder

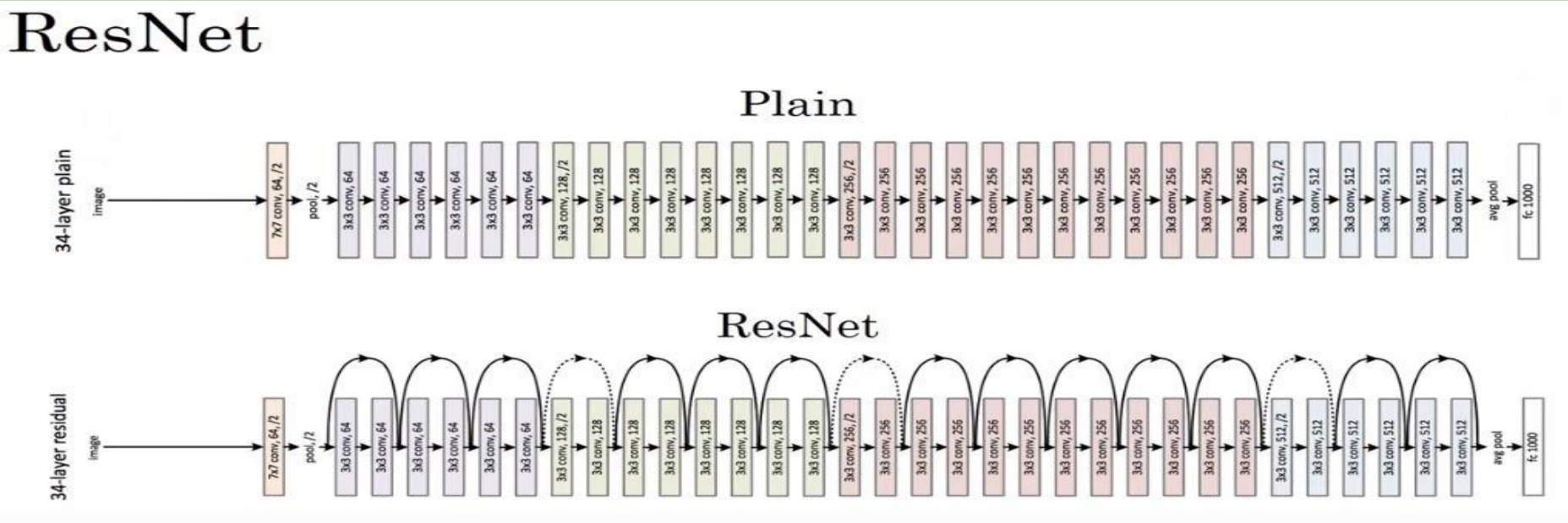




Reconstruction loss is measured against the original signal

Model Architecture

- **U_net architecture + Resnet34 in its encoder block**
  - Fast and computable deep conv network, owing to its residual blocks
  - Using its frozen layers (parameters) for transfer learning



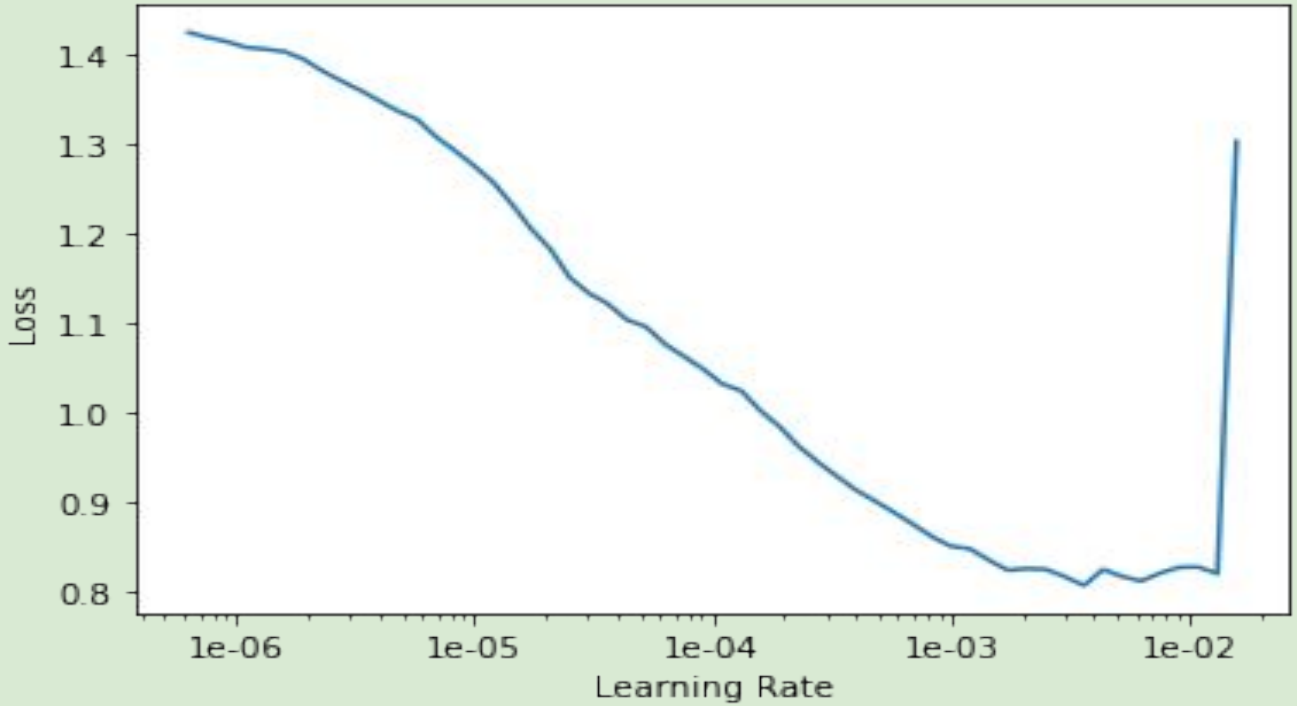
Data Transformations

- The following transforms applied to the training set :

1. flip_vert=True : flips images random by any possible 90 deg
2. max_rotate=10. : rotation of -10 to 10 deg randomly
3. max_zoom=1.05
4. max_warp=0.2 : accounting for variable perspectives in different images
5. max_lighting=0.1
6. cutout(n_holes=(1,4), length=(10, 160), p=0.75) : cuts out random boxes in the image, to account for possible water marks on the test images

Training stages:

1. Finding the optimum learning rate (for fastest loss descent) for frozen layers



Training stages

2. Training the model with Resnet layers frozen (i.e, non-trainable params)

At each epoch the params of the model with best performance on Val set are saved

3. The model is unfrozen
Learning rate is re-assessed
Model is re-trained with new LR and all trainable params

4. Best performing model (params + arch) is saved for future use

Two stages of transfer learning

1. Transfer learning from Resnet 34

    - During frozen state of resnet block
    - During unfrozen state, using the resnet's params as initial vals

2. Training the model with 400x400 size images to explore the performance
    - Saving the best epoch model, loading it back up, changing the size to 800x800, changing the Data bunch, applying it to the model (learn.data= newdata), then training the model with the new image sizes.

Challenges

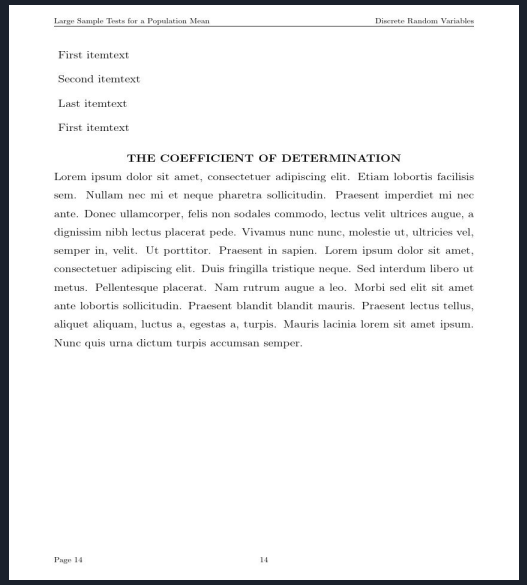
- **GPU memory outage, due to:**
  1. Application of multiple transforms
  2. High resolution input image
  3. Increasing batch size

- **Dataset style/format diversity**

- **Accuracy improvement for small objects (title, subtitle)**

Resolution Challenge for OCR

800x800 input size



First itemtext

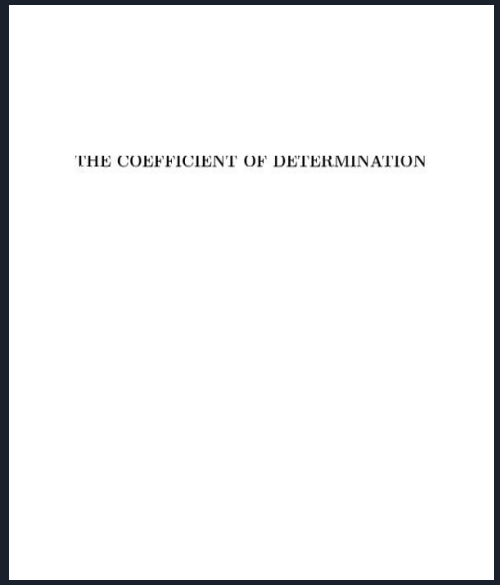
Second itemtext

Last itemtext

First itemtext

THE COEFFICIENT OF DETERMINATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

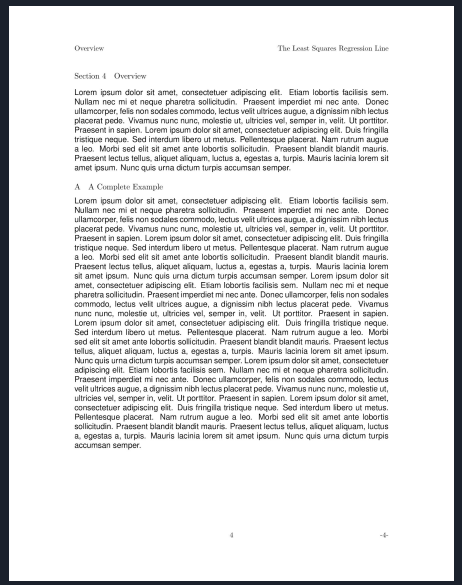


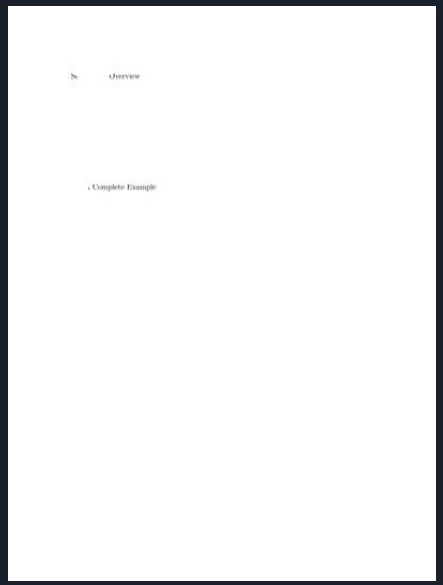
OCR

TXT

Resolution Challenge for OCR

400x400 input size



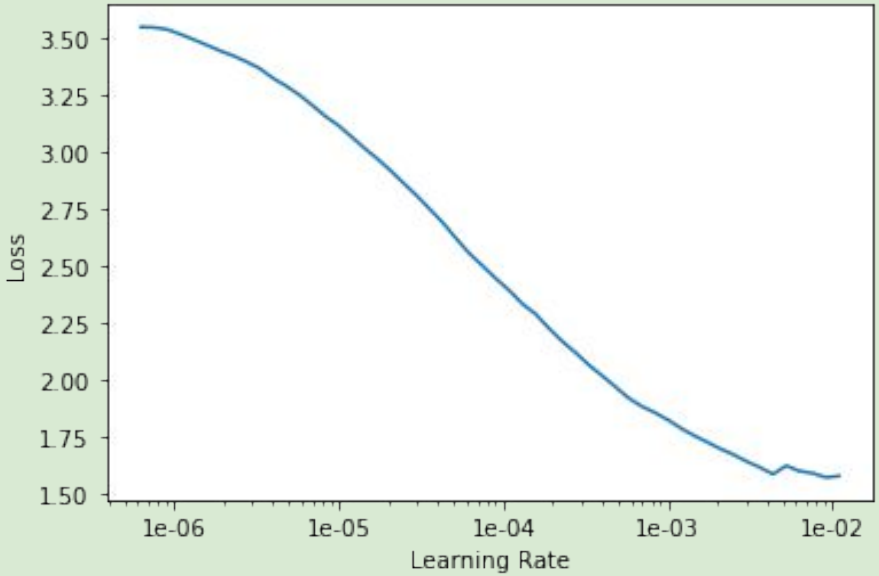


OCR

TXT

Learning Rate Assessments

LR before froze training



LR before unfroze training

Learning Rate importance

- Too high LR leads to too high val_loss regardless of # of epochs
- Too low lr leads to val_loss dropping very very slowly
- You want your LR to be large enough in the early stages of training so you don’t get stuck in a local minima/saddle points, but also low enough in the latter stages of training to stay in the global minima if it finds it
- Also, good LR can reduce the computation time by almost 3 to 5 times

Discriminative learning rates

- Using a range of learning rates to train the model's layers
- Training the early layers at slower rate than the latter layers
  i.e, how quickly the earliest layer of the model learns compared to the latest layer

Running into GPU memory outage

Parameters involved:

- Choosing the model, Resnet 18? Resnet 34? … different models perform different and also generate 31 to 41, to… million parameters
- Also freezing most of the layers or unfreezing all layers before training changes the number of parameters as well as the performance of the model
- Input image size (matters when training the model, and when trying to extract text, i.e, characters at the end)
- Dataset size, i.e, validation set size (matters to ensure there is a good distribution of all classes , poor presence of some classes in the images could lead to NAN values when reporting their individual accuracy values)
- Batch size (higher batch size generally leads to better training)
- Number and kind of transforms applied (matters when training the model to be reliable and precise in its prediction with any deviations in the test images from the training images
- Should I use an expensive model like resnet34 or 50 instead of 18 ? or I increase the size of my input image from 800X800 to higher values??

Trick to reduce GPU memory outage

Mixed_precision training

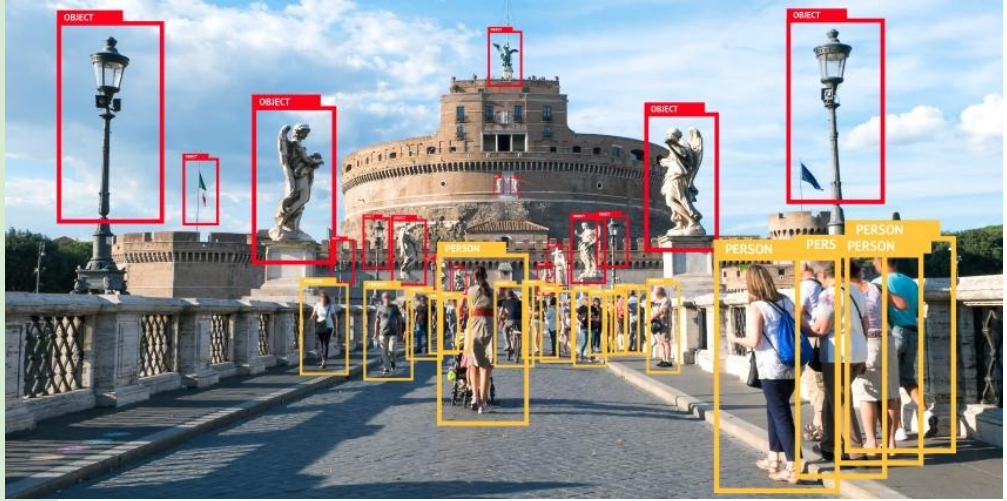
- Instead of using single-precision float numbers using half-precision floating numbers, 16 bit:
- ```
  learn = unet_learner(data, model, wd = wd, metrics = [Nobground_acc, Body_acc, List_acc, Header_Footer_acc , Title_acc]).to_fp16()
  ```
- Downside: requires very new GPU hardware
- Upside: can input higher images sizes, and/or higher bach sizes, faster, and possibly even better accuracies

This enabled me to go from 4 to 6 bach size with 800x800 images size with the following transforms on 5000 data size

```
get_transforms(
flip_vert=True, max_rotate=10.,
max_zoom=1.05,max_warp=0.2, max_lighting=0.1)
```
```

Another Case

Your client wants you to analyze whether they should advertise food after a travel video clip or not



Data:

No comments

No dialogue

Approach:

Create a time-series text dataset

Apply NLP

Data Cleaning Metric

1	file	pages	nerror	errormagr	nerrorPerPage
953	trainX/pdf_input_1148.pdf	1	0	0	0
954	trainX/pdf_input_1476.pdf	3	0	0	0
955	trainX/pdf_input_1681.pdf	2	0	0	0
956	trainX/pdf_input_0026.pdf	4	0	0	0
957	trainX/pdf_input_0018.pdf	27	0	0	0
958	trainX/pdf_input_0256.pdf	8	0	0	0
959	trainX/pdf_input_1018.pdf	5	1	0.24503	0.2
960	trainX/pdf_input_0342.pdf	23	1	1.51865	0.043478261
961	trainX/pdf_input_0450.pdf	9	1	0.87756	0.111111111
962	trainX/pdf_input_1007.pdf	5	1	0.57759	0.2

Image credits

https://dailynlp.com/wp-content/uploads/2012/05/confused.jpg

https://medium.com/@ksuryaremanan/beginners-guide-to-object-detection-algorithms-6620fb31c375

https://security-img.scdn6.secure.raxcdn.com/moreimages/AI-object-detection-850x425.jpg

https://miro.medium.com/max/539/1*rtHnQ41xi3MgliJWEw8wrA.jpeg

Resnet: Andrew NG https://www.youtube.com/watch?v=RYth6EbBUqM

Table 15: Confusion Matrix for a 3x3 classifier. Please see text for further explanation and its connection to the Classification Report in table (16).

	Class1 predicted	Class2 predicted	Class3 predicted
Class1 actual	A_{11}	A_{12}	A_{13}
Class2 actual	A_{21}	A_{22}	A_{23}
Class3 actual	A_{31}	A_{32}	A_{33}

Classification Report
Recall

Table (16) presents three different evaluation measures: Precision, Recall, and f1 Score.

Recall: For instance, test set contained S1[67] items in class 1, model predicted A_{11} of them correctly in class 1, therefore all Recalls are defined as[68]:
$R1 = A_{11}/S1$
$R2 = A_{22}/S2$
$R3 = A_{33}/S3$.

Classification Report
Precision

Precision: Now observe the first column of table (15), for instance. The classifier made $A_{11} + A_{21} + A_{31}$ number of predictions to be in class 1, of which, A_{11} of them belong to true class 1 (i.e., correct predictions), A_{21} of them belong to class 2 (i.e., misclassifications), and A_{31} of them belong to class 3 (i.e., misclassifications). Therefore, all Precisions would be defined as[69]:
$P1 = A_{11}/(A_{11} + A_{21} + A_{31})$
$P2 = A_{22}/(A_{12} + A_{22} + A_{32})$
$P3 = A_{33}/(A_{13} + A_{23} + A_{33})$

$$f1s1 = \frac{2 * P1 * R1}{P1 + R1}$$
$$f1s2 = \frac{2 * P2 * R2}{P2 + R2}$$
$$f1s3 = \frac{2 * P3 * R3}{P3 + R3}$$

Table 16: Classification report. Support values S1, S2, and S3 represent the number of samples from each class in the test set. For explanations of Precision, Recall, and f1-Score please see text.

Class	Precision	Recall	f1 Score	Support
1	P1	R1	f1s1	S1
2	P2	R2	f1s2	S2
3	P3	R3	f1s3	S3
Ave/Total	Ave P	Ave R	Ave fs	Total Support

Recall
Precision
F1Score