LayoutLmv3 (Data Augmentation and Imbalance mitigation)

Hardware Setup

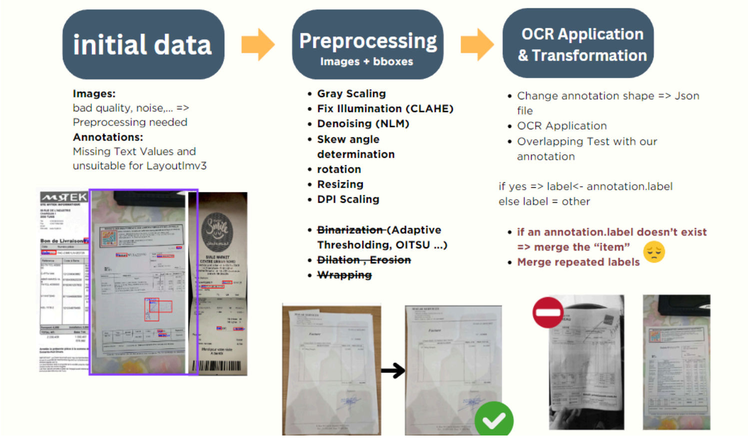
We conducted our experiments using an **NVIDIA GeForce RTX 3050 Laptop GPU with 4GB of dedicated memory.** The limited GPU memory was a key consideration in our experiment setup and model configuration.

Data Preparation

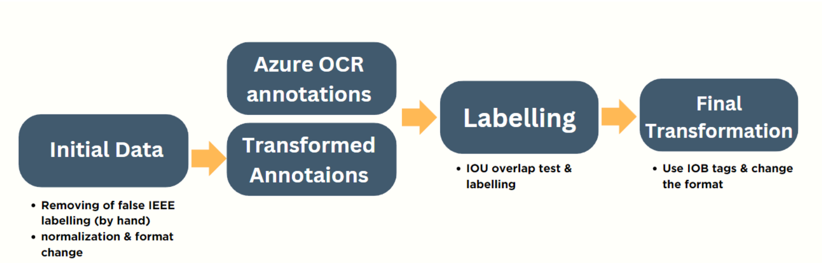
The dataset comprised **196 scanned invoices**, with annotations initially provided in CSV format and subsequently converted to JSON to align with model input requirements. Manual corrections were made for errors in the 'ieee' labels.

OCR Processing

1. **Pytesseract OCR:** Initial OCR was performed using Pytesseract, which required preprocessing of images to improve text recognition accuracy. Post-OCR, data was reformatted to fit the input requirements of LayoutLM **(image, words, ner_tags, bounding_boxes).**



2. Azure OCR: Due to suboptimal results from Pytesseract affecting model performance, Azure OCR was utilized to **ensure higher accuracy** in text extraction.
3. **Labeling via IOU**: The Intersection Over Union (IOU) metric was employed to align OCR results with existing labels, facilitating accurate training data preparation.



Data Imbalance

The dataset exhibits significant class imbalance, which is detailed as follows in the training set (80% of total data):

- 'O': 25,989
- 'B-title': 137
- 'B-date': 146
- 'B-ieee': 133
- 'B-total': 150
- 'B-totalValue': 147

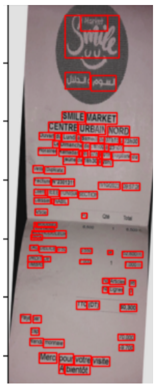
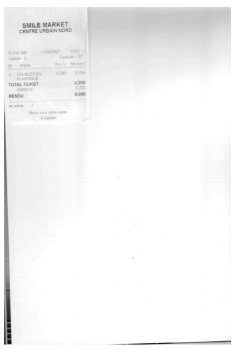
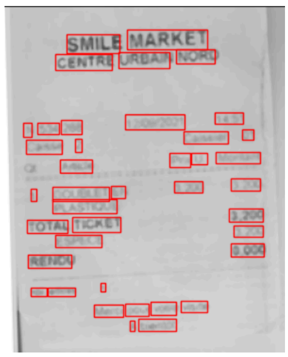
the labels needed to be in the IOB(inside, outside, beginning) form : **{'O', 'B-title', 'I-title', 'B-date', 'I-date', 'B-ieee', 'I-ieee', 'B-total','I-total', 'B-totalValue', 'I-totalValue'}**

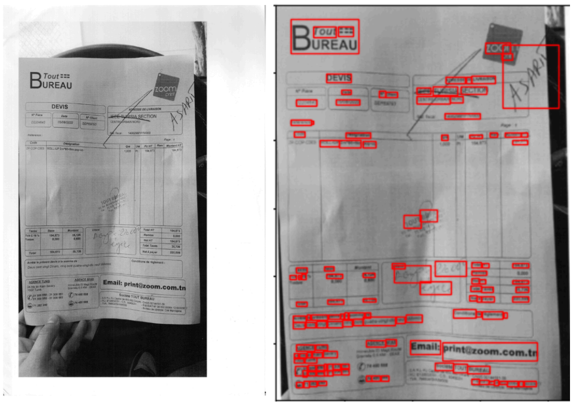
Data Augmentation

Even small rotations (limit=) caused cropping out essential textual elements, leading to mismatches between the words, labels, and their corresponding bounding boxes. To mitigate this issue, I first applied **padding** to the images before rotation. We then developed **a script to identify the minimum and maximum x and y coordinates across all bounding boxes**, ensuring that the cropping maintained **a margin of 20 pixels** from these boundaries. This approach not only **preserved crucial data** but also standardized image dimensions across the dataset, significantly enhancing model performance. Furthermore, the augmentation techniques **improved the quality of images with complex backgrounds that were initially challenging to preprocess**, ensuring more consistent and clear inputs for model training.

Steps:

- Median Blur, Color Jitter for image enhancement.
- Padding with a probability of 1
- Rotation: Limit of 5 degrees, applied with **inter cubic interpolation**
- Resizing using **bilinear sampling** to maintain uniform image dimensions, beneficial for consistent model performance.



Model Training and Fine-Tuning

To prevent overfitting, a **weight decay regularization of 0.01** is implemented.

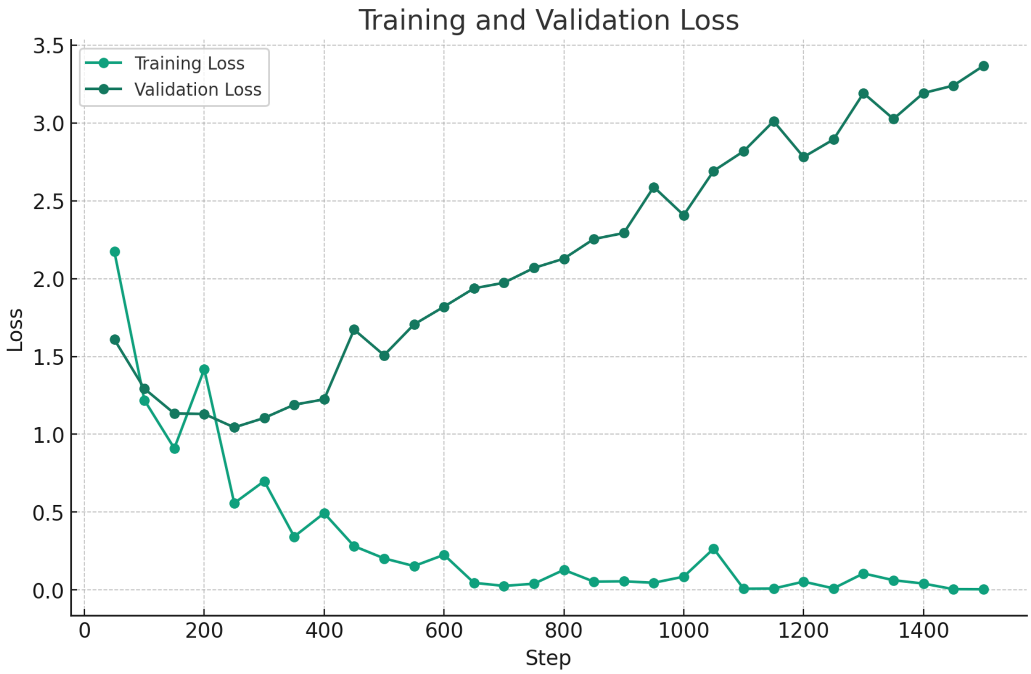
- **Loss Function:** Cross entropy with class weights and Focal loss were tested to address data imbalance.

	GFlops	Batch size	epoch	Learning rate	Inference time (seconds)	Params (Millions)	Training loss	Validation loss	F1 score	Accuracy	Recall	Precision
LMv3 (weighted Cross Entropy)	79.44	4	38	5e-5	1.809	125.93	0.10	3.19	0.50	0.96	0.52	0.48
LMv3 (Focal Loss)	44.83	4	50	5e-5	2.08	125.93	0.002	0.036	0.56	0.97	0.48	0.68

NB: (I tried using a batch size of 8 and LR= 10e-4 but the training always reached some point and then an error occurred, upon inspection i figured that the model wasn't "stable enough" and some forums advised of decreasing the lr/batch size)

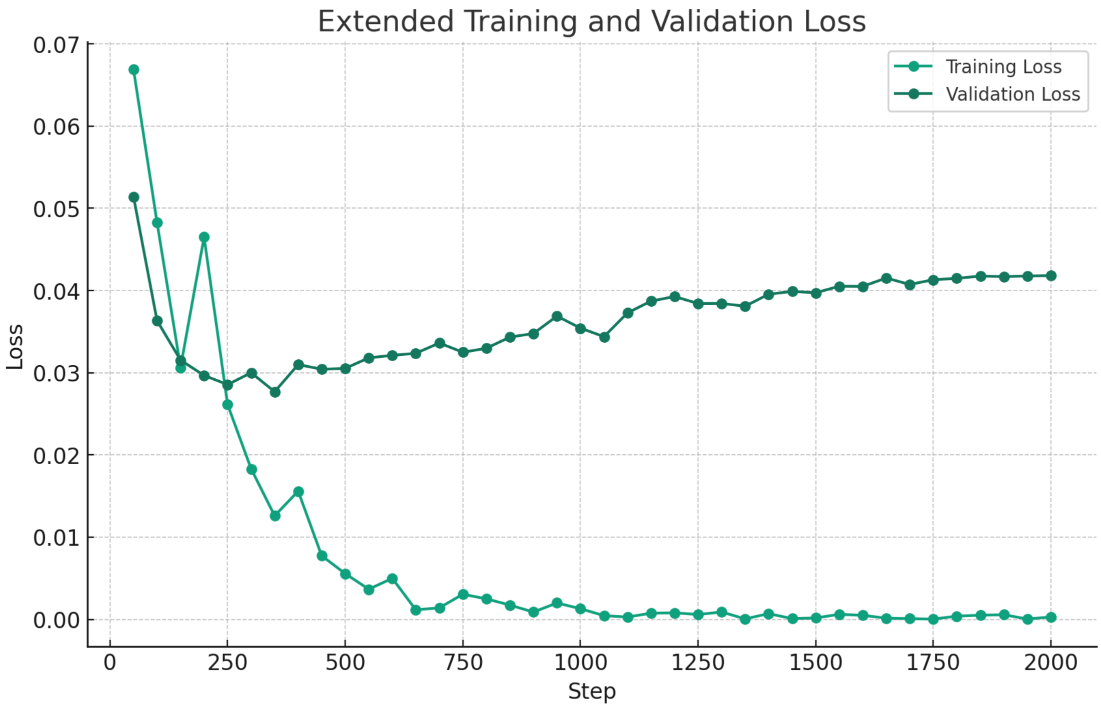
Training and validation loss plots & Inference :

Weighted Cross Entropy Loss: **(lasted 16h to 38 epochs even though I used GPU ⇒ Isuspect that the result is mistaken because of hardware issues)**

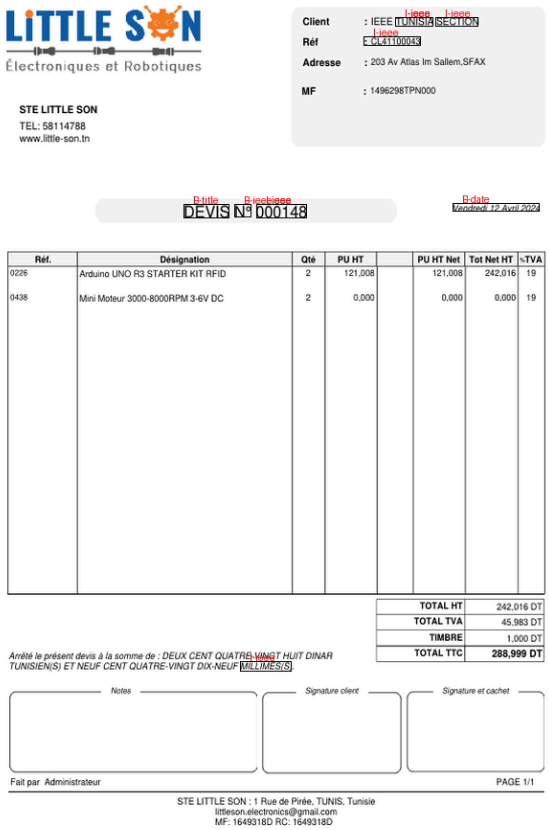


Inference: (shows only class other)

Focal Loss:



Inference: (plotting all labels besides other)



LiTTLE SON
Électroniques et Robotiques

STE LITTLE SON
TEL: 58114788
www.little-son.tn

Client	: IEEE TUNISIA SECTION
Réf	: CL41100043
Adresse	: 203 Av Atlas Im Sallem,SFAX
MF	: 1496298TPN000

DEVIS N° 000148

Vendredi 12 Avril 2024

Réf.	Désignation	Qté	PU HT		PU HT Net	Tot Net HT	%TVA
0226	Arduino UNO R3 STARTER KIT RFID	2	121,008		121,008	242,016	19
0438	Mini Moteur 3000-8000RPM 3-6V DC	2	0,000		0,000	0,000	19

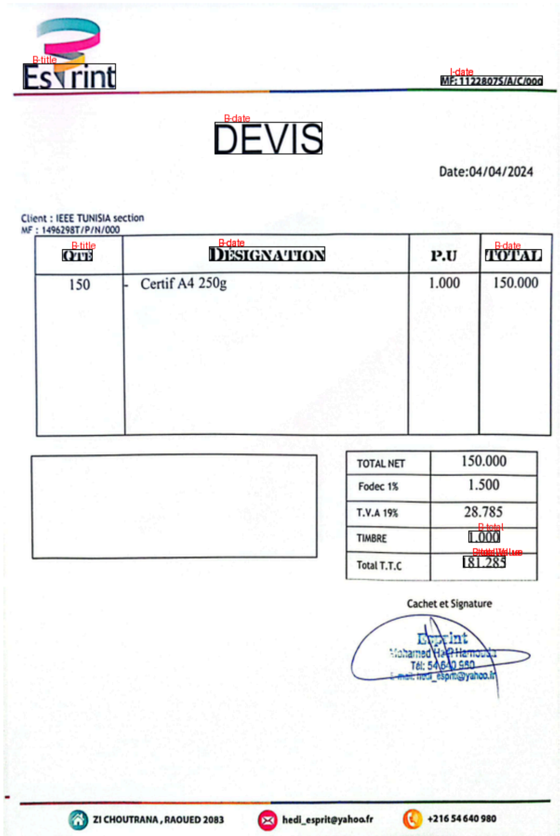
TOTAL HT	242,016 DT
TOTAL TVA	45,983 DT
TIMBRE	1,000 DT
TOTAL TTC	288,999 DT

Arrêté le présent devis à la somme de : DEUX CENT QUATRE-VINGT HUIT DINAR TUNISIEN(S) ET NEUF CENT QUATRE-VINGT DIX-NEUF MILLIMES(S).

Notes — Signature client — Signature et cachet

Fait par Administrateur — PAGE 1/1

STE LITTLE SON : 1 Rue de Pirée, TUNIS, Tunisie
littleson.electronics@gmail.com
MF: 1649318D RC: 1649318D



Esprint

MF: 1122807S/A/C/000

DEVIS

Date:04/04/2024

Client : IEEE TUNISIA section
MF : 1496298T/P/N/000

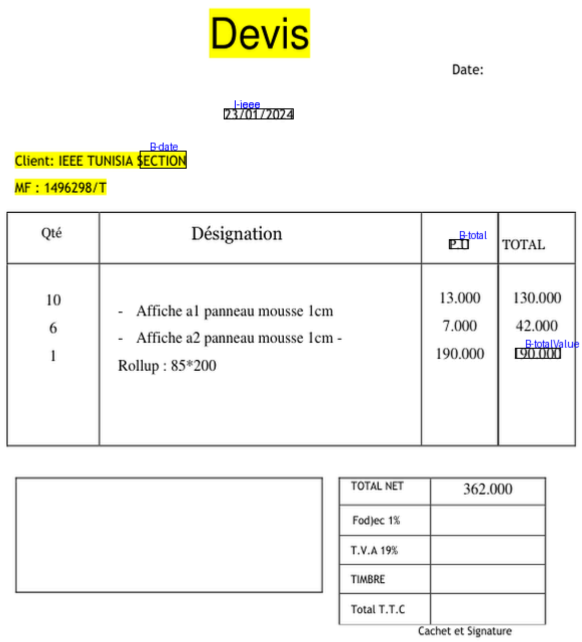
QTE	DESIGNATION	P.U	TOTAL
150	- Certif A4 250g	1.000	150.000

TOTAL NET	150.000
Fodec 1%	1.500
T.V.A 19%	28.785
TIMBRE	1.000
Total T.T.C	181.285

Cachet et Signature

Esprint
Mohamed Hedi Hamouda
Tél: 54 640 980
E-mail: hedi_esprit@yahoo.fr

ZI CHOUTRANA , RAOUED 2083 — hedi_esprit@yahoo.fr — +216 54 640 980



Devis

Date:

23/01/2024

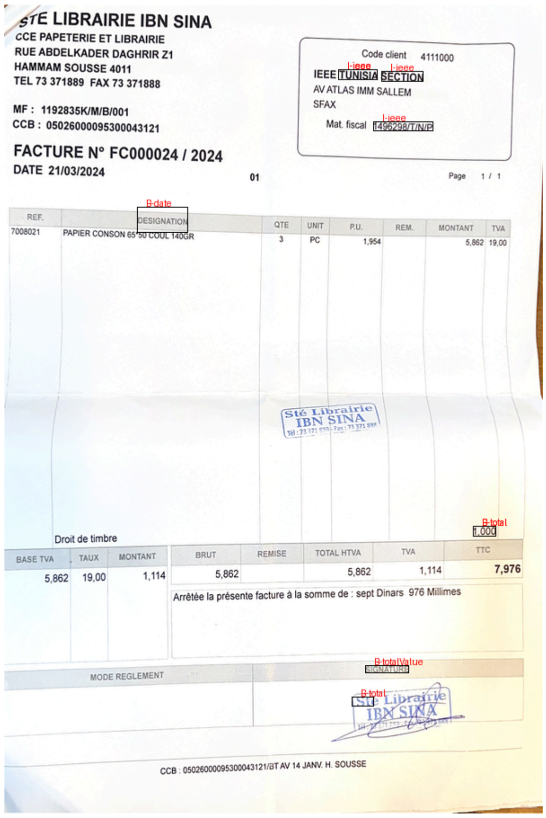
Client: IEEE TUNISIA SECTION

MF : 1496298/T

Qté	Désignation	P.U	TOTAL
10 6 1	- Affiche a1 panneau mousse 1cm - Affiche a2 panneau mousse 1cm - Rollup : 85*200	13.000 7.000 190.000	130.000 42.000 190.000

TOTAL NET	362.000
Fodjec 1%	
T.V.A 19%	
TIMBRE	
Total T.T.C	

Cachet et Signature



STE LIBRAIRIE IBN SINA
CCE PAPETERIE ET LIBRAIRIE
RUE ABDELKADER DAGHRIR Z1
HAMMAM SOUSSE 4011
TEL 73 371889 FAX 73 371888

MF : 1192835K/M/B/001
CCB : 05026000095300043121

Code client 4111000
IEEE TUNISIA SECTION
AV ATLAS IMM SALLEM
SFAX
Mat. fiscal 1496298/T/N/P

FACTURE N° FC000024 / 2024
DATE 21/03/2024

01

Page 1 / 1

REF.	DESIGNATION	QTE	UNIT	P.U.	REM.	MONTANT	TVA
7008021	PAPIER CONSON 65*50 COUL 140GR	3	PC	1,954		5,862	19,00

Droit de timbre 1,000

BASE TVA	TAUX	MONTANT	BRUT	REMISE	TOTAL HTVA	TVA	TTC
5,862	19,00	1,114	5,862		5,862	1,114	7,976

Arrêtée la présente facture à la somme de : sept Dinars 976 Millimes

MODE REGLEMENT

SIGNATURE

Sté Librairie IBN SINA

CCB : 05026000095300043121/BT AV 14 JANV. H. SOUSSE

Conclusion:

Cross entropy is the original loss function of the model but adding the **class weights** makes the process harder and way more resource demanding without any effects. **Focal Loss** shows better performance and is promising. training for more epochs and with more data might give some good results (worth experimenting with the handwritten and machine written data merged)

Next, Fedi and akram will each run the same code but train with LiLT and LayoutLMv2 so we can compare (OCR=False– since it's widely used and the model takes so much time to train, I suggest that we prioritize this comparison between models without using the built in OCR) – Run on colab