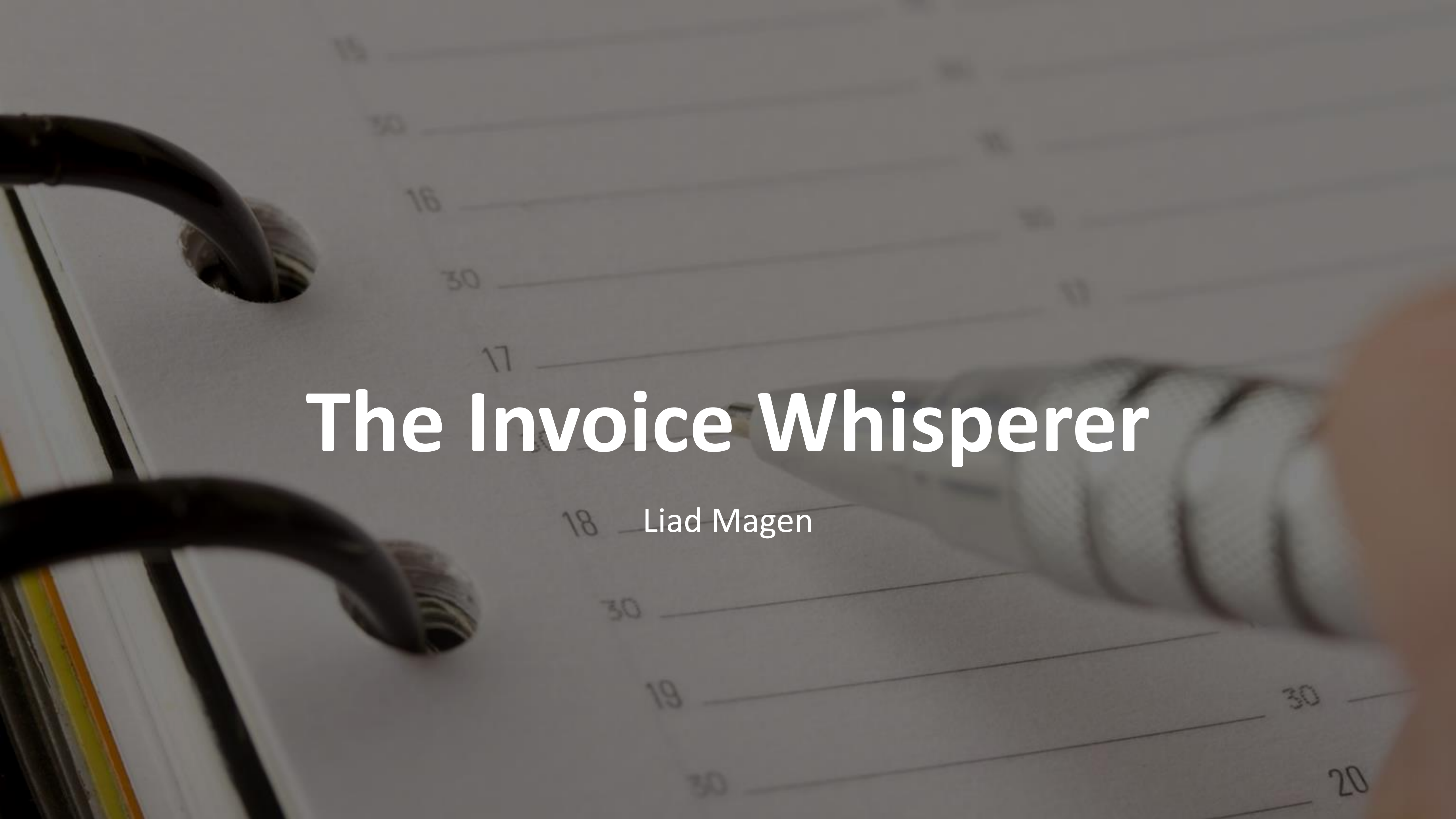
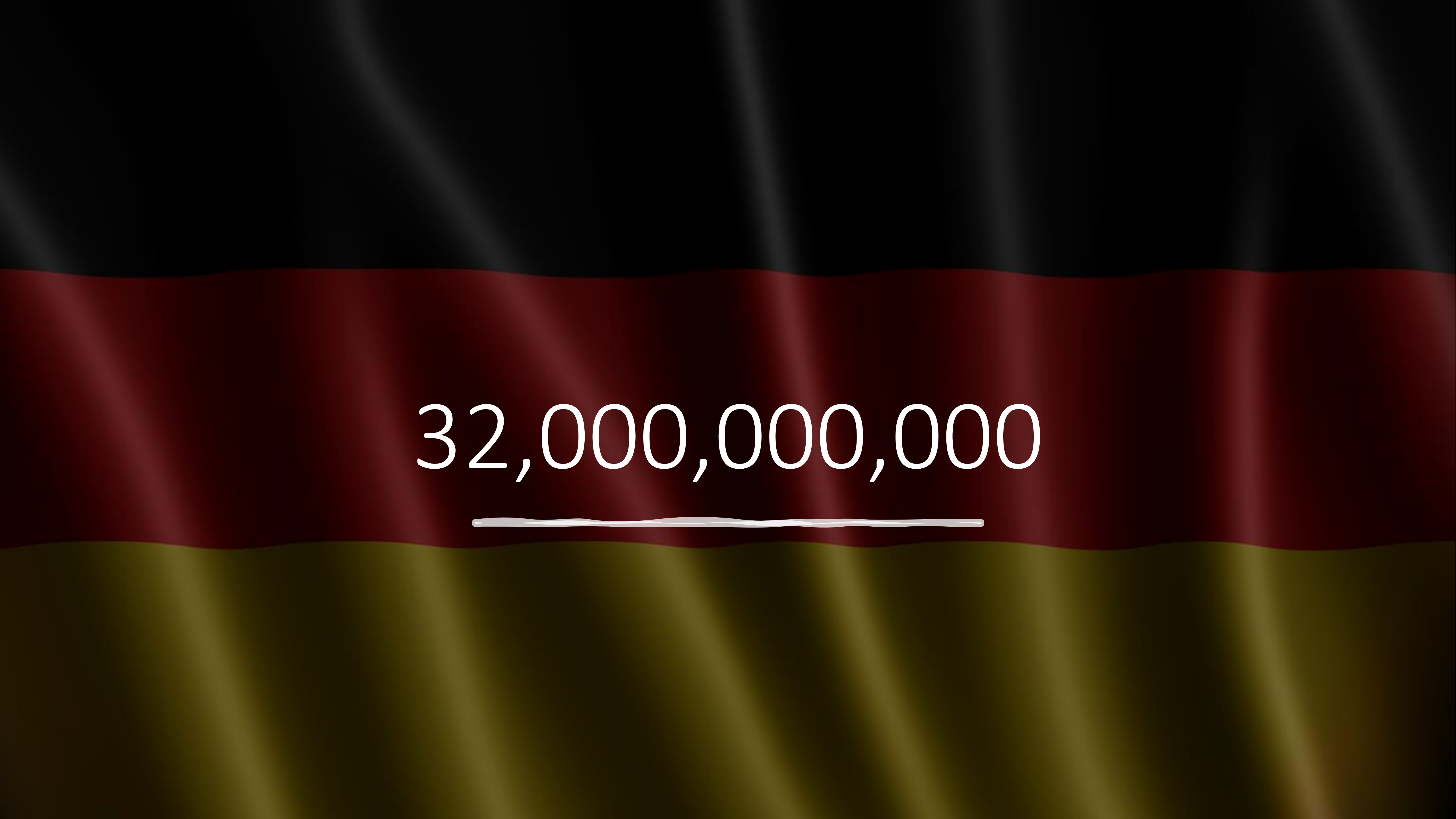


The Invoice Whisperer

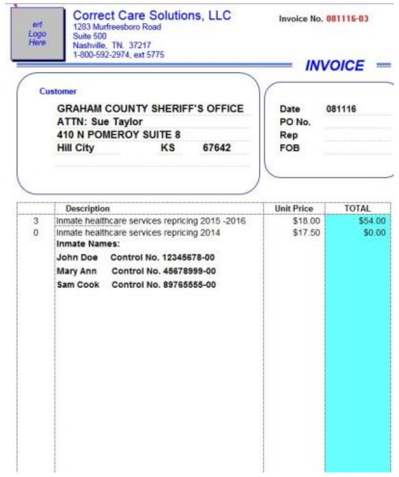
Liad Magen

919.5

1015



32,000,000,000



ert
Logo
Here

Correct Care Solutions, LLC
1283 Murfreesboro Road
Suite 500
Nashville, TN. 37217
1-800-592-2974, ext 5775

Invoice No. 081116-03

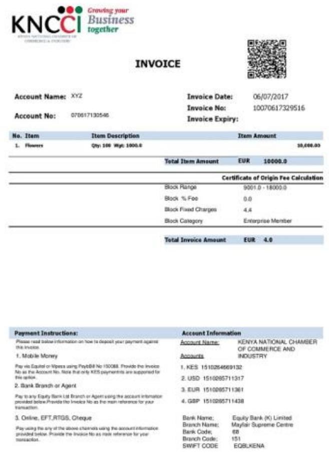
INVOICE

Customer

GRAHAM COUNTY SHERIFF'S OFFICE
ATTN: Sue Taylor
410 N POMEROY SUITE 8
Hill City KS 67642

Date	081116
PO No.	
Rep	
FOB	

	Description	Unit Price	TOTAL
3	Inmate healthcare services repricing 2015 -2016	$18.00	$54.00
0	Inmate healthcare services repricing 2014	$17.50	$0.00
	Inmate Names:		
	John Doe Control No. 12345678-00		
	Mary Ann Control No. 45678999-00		
	Sam Cook Control No. 89765555-00		



KNCCI Growing your Business together

INVOICE

Account Name:	XYZ	Invoice Date:	06/07/2017
Account No:	070617130546	Invoice No:	10070617329516
		Invoice Expiry:	

No.	Item	Item Description	Item Amount
1.	Flowers	Qty: 100 Wgt: 1000.0	10,000.00

Total Item Amount EUR 10000.0

Certificate of Origin Fee Calculation

Block Range	9001.0 - 18000.0
Block % Fee	0.0
Block Fixed Charges	4.4
Block Category	Enterprise Member

Total Invoice Amount EUR 4.0

Payment Instructions:

Please read below information on how to deposit your payment against this invoice.

1. Mobile Money

Pay via Equitel or Mpesa using PayBill No 150088. Provide the Invoice No as the Account No. Note that only KES paymentnts are supported for this option.

2. Bank Branch or Agent

Pay to any Equity Bank Ltd Branch or Agent using the account information provided below.Provide the Invoice No as the main reference for your transaction.

3. Online, EFT,RTGS, Cheque

Pay using the any of the above channels using the account information provided below. Provide the Invoice No as main reference for your transaction.

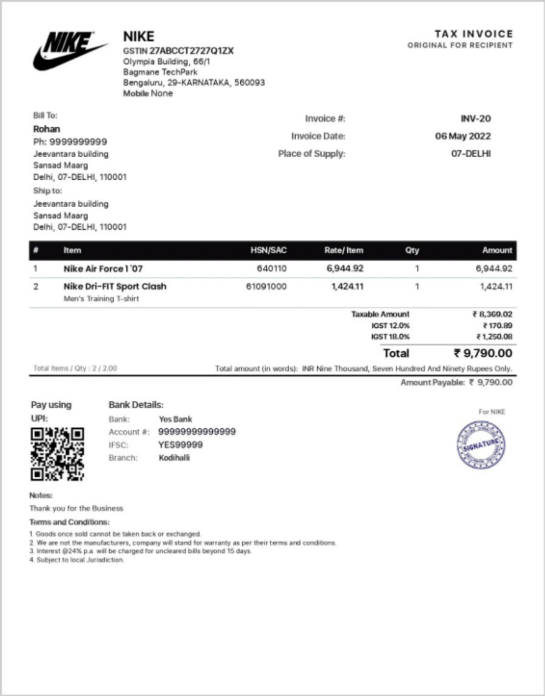
Account Information

Account Name: KENYA NATIONAL CHAMBER OF COMMERCE AND INDUSTRY

Accounts

1. KES 1510284669132
2. USD 1510285711317
3. EUR 1510285711361
4. GBP 1510285711438

Bank Name:	Equity Bank (K) Limited
Branch Name:	Mayfair Supreme Centre
Bank Code:	68
Branch Code:	151
SWIFT CODE	EQBLKENA



NIKE
GSTIN 27ABCCT2727Q1ZX
Olympia Building, 66/1
Bagmane TechPark
Bengaluru, 29-KARNATAKA, 560093
Mobile None

TAX INVOICE
ORIGINAL FOR RECIPIENT

Bill To:
Rohan
Ph: 9999999999
Jeevantara building
Sansad Maarg
Delhi, 07-DELHI, 110001

Ship to:
Jeevantara building
Sansad Maarg
Delhi, 07-DELHI, 110001

Invoice #:	INV-20
Invoice Date:	06 May 2022
Place of Supply:	07-DELHI

#	Item	HSN/SAC	Rate/ Item	Qty	Amount
1	Nike Air Force 1 '07	640110	6,944.92	1	6,944.92
2	Nike Dri-FIT Sport Clash Men's Training T-shirt	61091000	1,424.11	1	1,424.11

Taxable Amount	₹ 8,369.02
IGST 12.0%	₹ 170.89
IGST 18.0%	₹ 1,250.08
Total	**₹ 9,790.00**

Total Items / Qty : 2 / 2.00

Total amount (in words): INR Nine Thousand, Seven Hundred And Ninety Rupees Only.

Amount Payable: ₹ 9,790.00

Pay using UPI:

Bank Details:

Bank:	Yes Bank
Account #:	99999999999999
IFSC:	YES99999
Branch:	Kodihalli

For NIKE

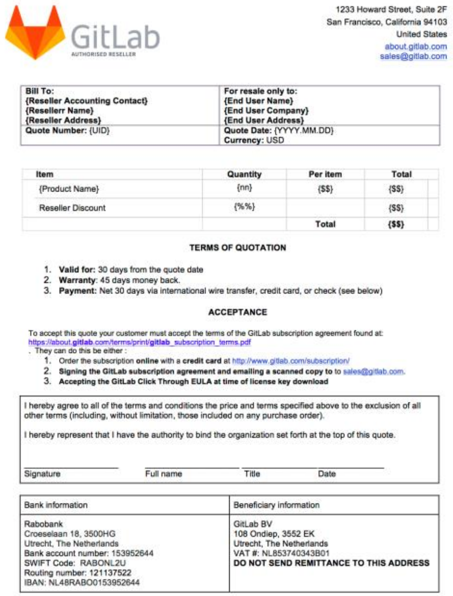
SIGNATURE

Notes:

Thank you for the Business

Terms and Conditions:

1. Goods once sold cannot be taken back or exchanged.
2. We are not the manufacturers, company will stand for warranty as per their terms and conditions.
3. Interest @24% p.a. will be charged for uncleared bills beyond 15 days.
4. Subject to local Jurisdiction.



GitLab
AUTHORISED RESELLER

1233 Howard Street, Suite 2F
San Francisco, California 94103
United States
about.gitlab.com
sales@gitlab.com

Bill To:
{Reseller Accounting Contact}
{Resellerr Name}
{Reseller Address}
	For resale only to:
	{End User Name}
	{End User Company}
	{End User Address}

Quote Number: {UID}	**Quote Date:** {YYYY.MM.DD} **Currency:** USD

Item	Quantity	Per item	Total
{Product Name}	{nn}	{$$}	{$$}
Reseller Discount	{%%}		{$$}
		Total	**{$$}**

TERMS OF QUOTATION

1. **Valid for:** 30 days from the quote date
2. **Warranty**: 45 days money back.
3. **Payment:** Net 30 days via international wire transfer, credit card, or check (see below)

ACCEPTANCE

To accept this quote your customer must accept the terms of the GitLab subscription agreement found at: https://about.gitlab.com/terms/print/gitlab_subscription_terms.pdf
. They can do this be either :

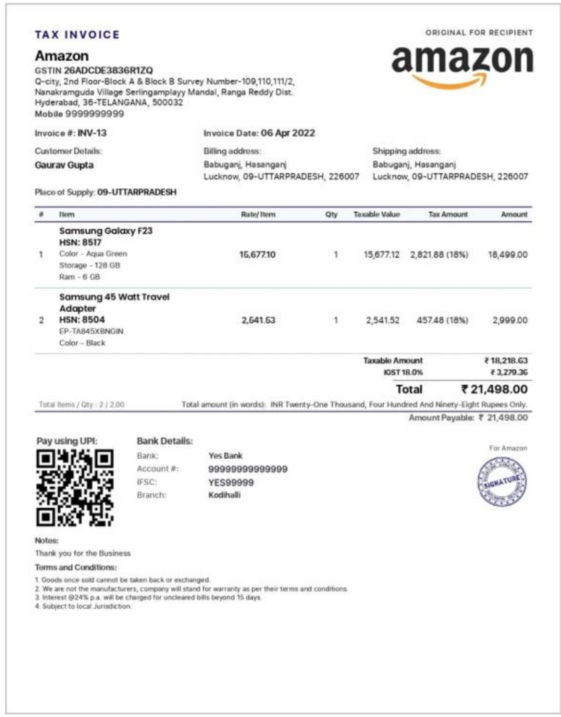
1. Order the subscription **online** with a **credit card** at http://www.gitlab.com/subscription/
2. **Signing the GitLab subscription agreement and emailing a scanned copy to** to sales@gitlab.com.
3. **Accepting the GitLab Click Through EULA at time of license key download**

I hereby agree to all of the terms and conditions the price and terms specified above to the exclusion of all other terms (including, without limitation, those included on any purchase order).

I hereby represent that I have the authority to bind the organization set forth at the top of this quote.

Signature Full name Title Date

Bank information	Beneficiary information
Rabobank Croeselaan 18, 3500HG Utrecht, The Netherlands Bank account number: 153952644 SWIFT Code: RABONL2U Routing number: 121137522 IBAN: NL48RABO0153952644	GitLab BV 108 Ondiep, 3552 EK Utrecht, The Netherlands VAT #: NL853740343B01 **DO NOT SEND REMITTANCE TO THIS ADDRESS**



TAX INVOICE
ORIGINAL FOR RECIPIENT

amazon

Amazon
GSTIN 26ADCDE3836R1ZQ
Q-city, 2nd Floor-Block A & Block B Survey Number-109,110,111/2,
Nanakramguda Village Serlingamplayy Mandal, Ranga Reddy Dist.
Hyderabad, 36-TELANGANA, 500032
Mobile 9999999999

Invoice #: **INV-13**

Invoice Date: **06 Apr 2022**

Customer Details:
Gaurav Gupta

Billing address:
Babuganj, Hasanganj
Lucknow, 09-UTTARPRADESH, 226007

Shipping address:
Babuganj, Hasanganj
Lucknow, 09-UTTARPRADESH, 226007

Place of Supply: **09-UTTARPRADESH**

#	Item	Rate/ Item	Qty	Taxable Value	Tax Amount	Amount
1	**Samsung Galaxy F23** **HSN: 8517** Color - Aqua Green Storage - 128 GB Ram - 6 GB	15,677.10	1	15,677.12	2,821.88 (18%)	18,499.00
2	**Samsung 45 Watt Travel Adapter** **HSN: 8504** EP-TA845XBNGIN Color - Black	2,641.53	1	2,541.52	457.48 (18%)	2,999.00

Taxable Amount	₹ 18,218.63
IGST 18.0%	₹ 3,279.36
Total	**₹ 21,498.00**

Total Items / Qty : 2 / 2.00

Total amount (in words): INR Twenty-One Thousand, Four Hundred And Ninety-Eight Rupees Only.

Amount Payable: ₹ 21,498.00

Pay using UPI:

Bank Details:

Bank:	Yes Bank
Account #:	99999999999999
IFSC:	YES99999
Branch:	Kodihalli

For Amazon

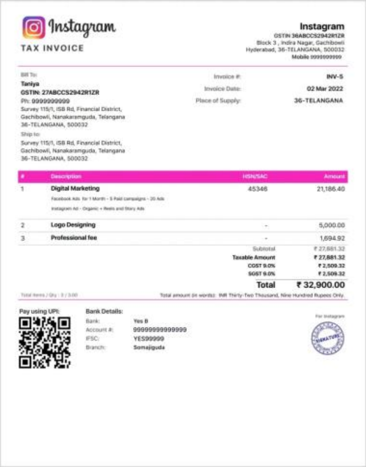
SIGNATURE

Notes:

Thank you for the Business

Terms and Conditions:

1. Goods once sold cannot be taken back or exchanged.
2. We are not the manufacturers, company will stand for warranty as per their terms and conditions.
3. Interest @24% p.a. will be charged for uncleared bills beyond 15 days.
4. Subject to local Jurisdiction.



Instagram
TAX INVOICE

Instagram
GSTIN 36ABCCS2942R1ZR
Block 3 , Indira Nagar, Gachibowli
Hyderabad, 36-TELANGANA, 500032
Mobile 9999999999

Bill To:
Taniya
GSTIN: 27ABCCS2942R1ZR
Ph: 9999999999
Survey 115/1, ISB Rd, Financial District,
Gachibowli, Nanakaramguda, Telangana
36-TELANGANA, 500032

Ship to:
Survey 115/1, ISB Rd, Financial District,
Gachibowli, Nanakaramguda, Telangana
36-TELANGANA, 500032

Invoice #:	INV-5
Invoice Date:	02 Mar 2022
Place of Supply:	36-TELANGANA

#	Description	HSN/SAC	Amount
1	**Digital Marketing** Facebook Ads for 1 Month - 5 Paid campaigns - 20 Ads Instagram Ad - Organic + Reels and Story Ads	45346	21,186.40
2	**Logo Designing**	-	5,000.00
3	**Professional fee**	-	1,694.92

Subtotal	₹ 27,881.32
Taxable Amount	₹ 27,881.32
CGST 9.0%	₹ 2,509.32
SGST 9.0%	₹ 2,509.32
Total	**₹ 32,900.00**

Total Items / Qty : 3 / 3.00

Total amount (in words): INR Thirty-Two Thousand, Nine Hundred Rupees Only.

Pay using UPI:

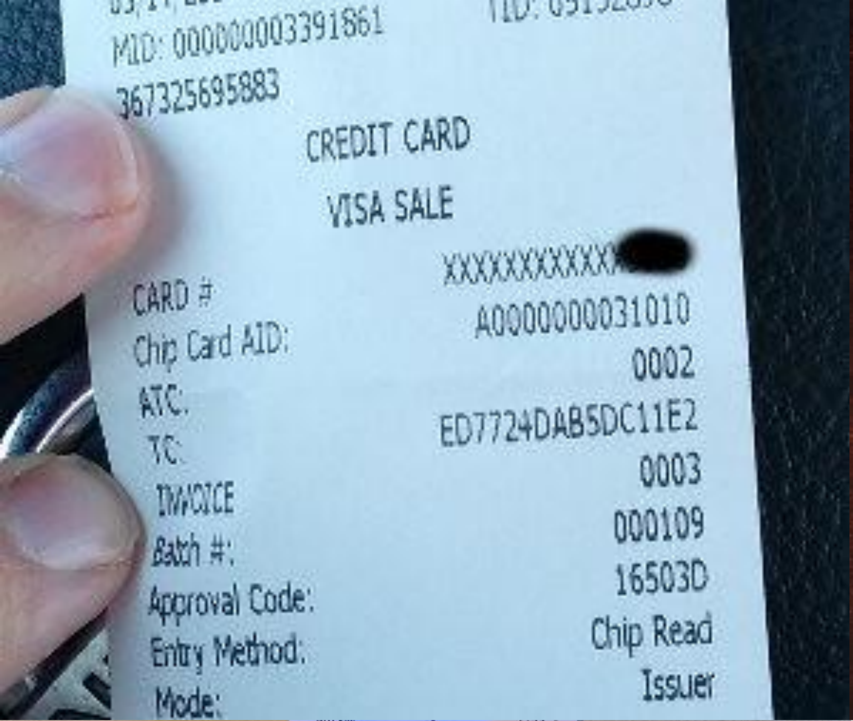
Bank Details:

Bank:	Yes B
Account #:	99999999999999
IFSC:	YES99999
Branch:	Somajiguda

For Instagram

SIGNATURE

Diversity

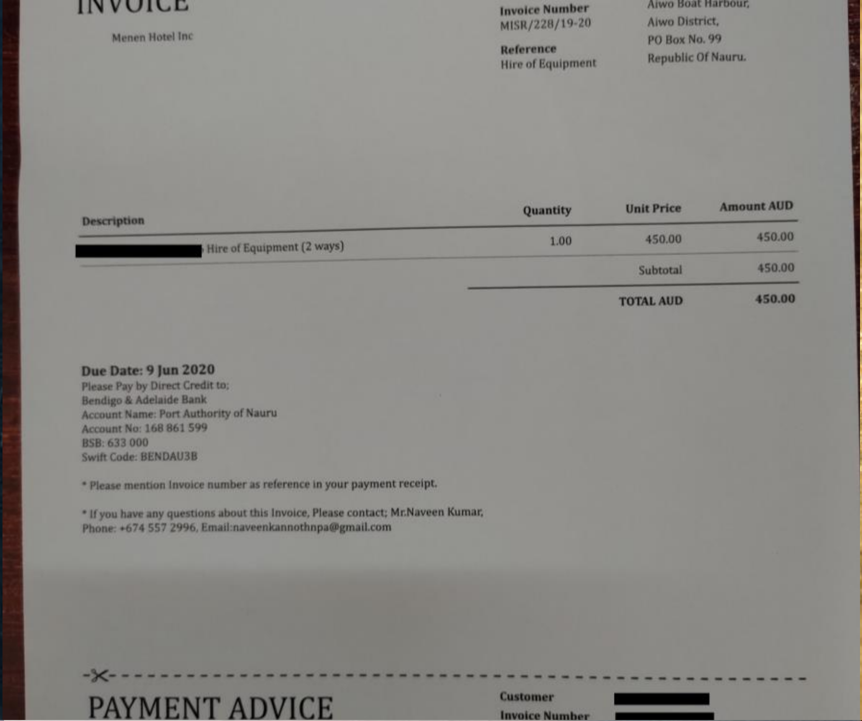


MID: 000000003391861
367325695883

CREDIT CARD

VISA SALE

CARD #	XXXXXXXXXXXX[redacted]
Chip Card AID:	A0000000031010
ATC:	0002
TC:	ED7724DAB5DC11E2
INVOICE	0003
Batch #:	000109
Approval Code:	16503D
Entry Method:	Chip Read
Mode:	Issuer



INVOICE

Menen Hotel Inc

Invoice Number
MISR/228/19-20

Reference
Hire of Equipment

Aiwo Boat Harbour,
Aiwo District,
PO Box No. 99
Republic Of Nauru.

Description	Quantity	Unit Price	Amount AUD
[redacted] Hire of Equipment (2 ways)	1.00	450.00	450.00
		Subtotal	450.00
		TOTAL AUD	**450.00**

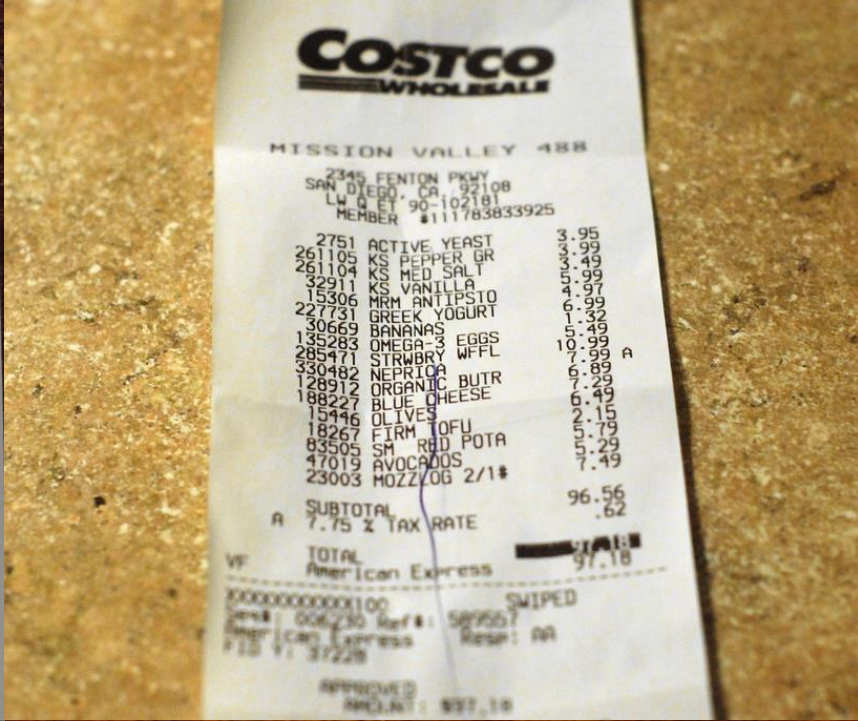
Due Date: 9 Jun 2020
Please Pay by Direct Credit to;
Bendigo & Adelaide Bank
Account Name: Port Authority of Nauru
Account No: 168 861 599
BSB: 633 000
Swift Code: BENDAU3B

* Please mention Invoice number as reference in your payment receipt.

* If you have any questions about this Invoice, Please contact; Mr.Naveen Kumar, Phone: +674 557 2996, Email:naveenkannothnpa@gmail.com

PAYMENT ADVICE

Customer [redacted]
Invoice Number [redacted]



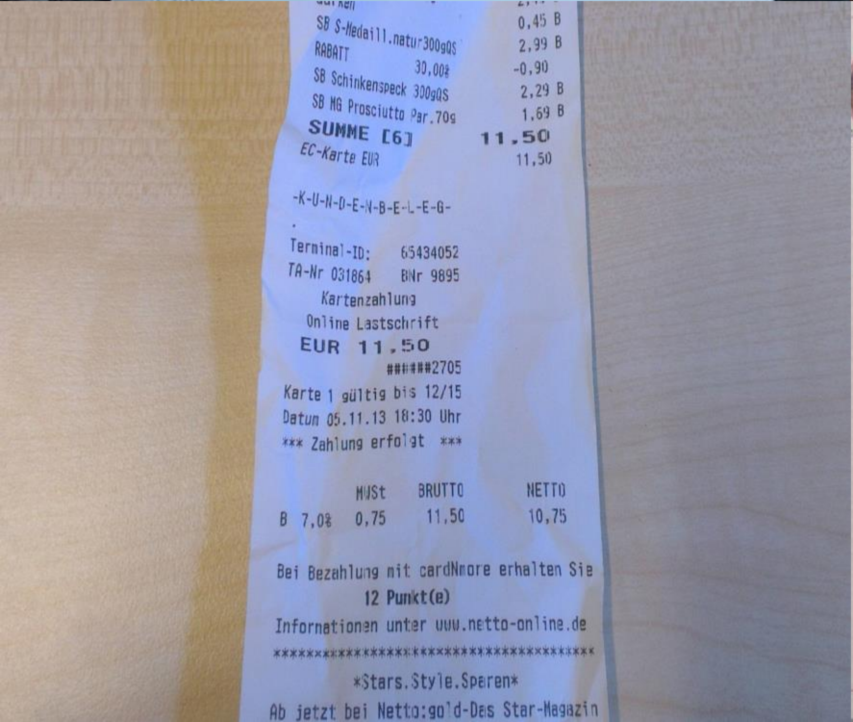
COSTCO WHOLESALE

MISSION VALLEY 488
2345 FENTON PKWY
SAN DIEGO, CA 92108
LW Q ET 90-102181
MEMBER #111783833925

2751	ACTIVE YEAST	3.95	
261105	KS PEPPER GR	3.99	
261104	KS MED SALT	3.49	
32911	KS VANILLA	5.99	
15306	MRM ANTIPSTO	4.97	
227731	GREEK YOGURT	6.99	
30669	BANANAS	1.32	
135283	OMEGA-3 EGGS	5.49	
285471	STRWBRY WFFL	10.99	
330482	NEPRICA	7.99	A
128912	ORGANIC BUTR	6.89	
188227	BLUE CHEESE	7.29	
15446	OLIVES	6.49	
18267	FIRM TOFU	2.15	
83505	SM RED POTA	5.79	
47019	AVOCADOS	5.29	
23003	MOZZLOG 2/1$	7.49	
	SUBTOTAL	96.56	
A	7.75 % TAX RATE	.62	
	TOTAL	[redacted]	
VF	American Express	97.18	

XXXXXXXXXXXX100 SWIPED
Seq#: 006230 Ref#: 589557
American Express Resp: AA
PID #: 37228

APPROVED
AMOUNT: $97.18



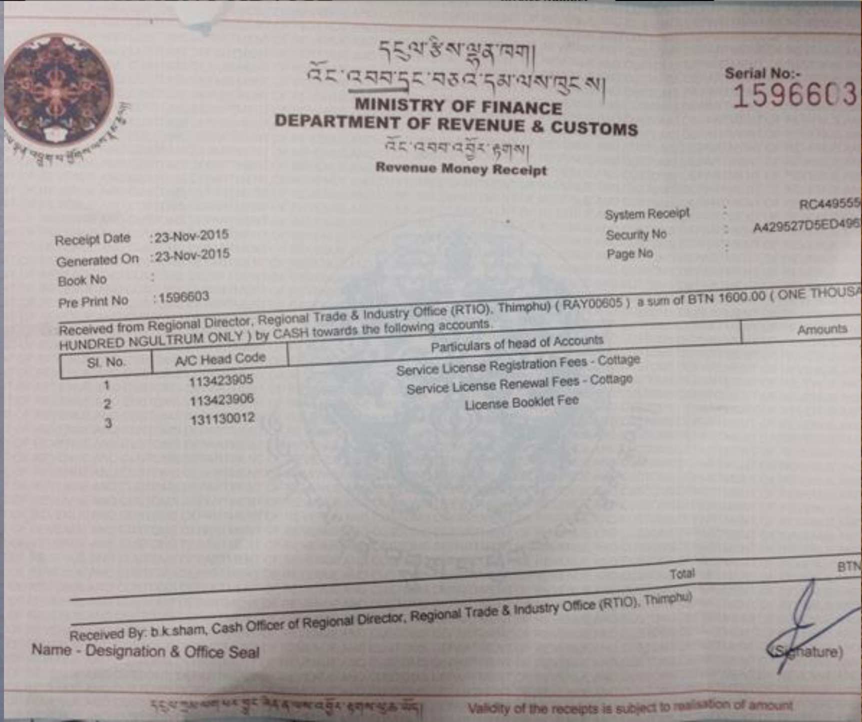
SB S-Medaill.natur300gQS	0,45 B
RABATT	2,99 B
30,00%	-0,90
SB Schinkenspeck 300gQS	2,29 B
SB MG Prosciutto Par.70g	1,69 B
SUMME [6]	**11,50**
EC-Karte EUR	11,50

-K-U-N-D-E-N-B-E-L-E-G-

Terminal-ID: 65434052
TA-Nr 031864 BNr 9895
Kartenzahlung
Online Lastschrift
EUR 11.50
######2705
Karte 1 gültig bis 12/15
Datum 05.11.13 18:30 Uhr
*** Zahlung erfolgt ***

	MWSt	BRUTTO	NETTO
B 7,0%	0,75	11,50	10,75

Bei Bezahlung mit cardNmore erhalten Sie
12 Punkt(e)
Informationen unter www.netto-online.de
****
Stars.Style.Sparen
Ab jetzt bei Netto:gold-Das Star-Magazin



དངུལ་རྩིས་ལྷན་ཁག།
འོང་འབབ་དང་བཅའ་དམ་ལས་ཁུངས།
MINISTRY OF FINANCE
DEPARTMENT OF REVENUE & CUSTOMS
འོང་འབབ་འཛིན་ཤོག།
Revenue Money Receipt

Serial No:-
1596603

Receipt Date :23-Nov-2015
Generated On :23-Nov-2015
Book No :
Pre Print No :1596603

System Receipt : RC449555
Security No : A429527D5ED496
Page No :

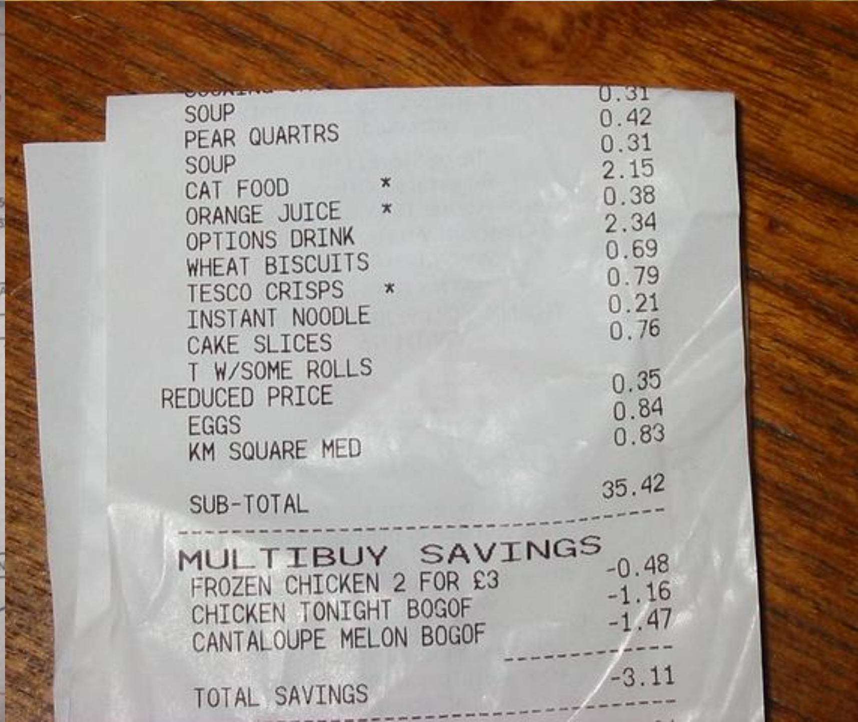
Received from Regional Director, Regional Trade & Industry Office (RTIO), Thimphu) (RAY00605) a sum of BTN 1600.00 (ONE THOUSA HUNDRED NGULTRUM ONLY) by CASH towards the following accounts.

Sl. No.	A/C Head Code	Particulars of head of Accounts	Amounts
1	113423905	Service License Registration Fees - Cottage	
2	113423906	Service License Renewal Fees - Cottage	
3	131130012	License Booklet Fee	
		Total	BTN

Received By: b.k.sham, Cash Officer of Regional Director, Regional Trade & Industry Office (RTIO), Thimphu
Name - Designation & Office Seal

(Signature)

[illegible] Validity of the receipts is subject to realisation of amount

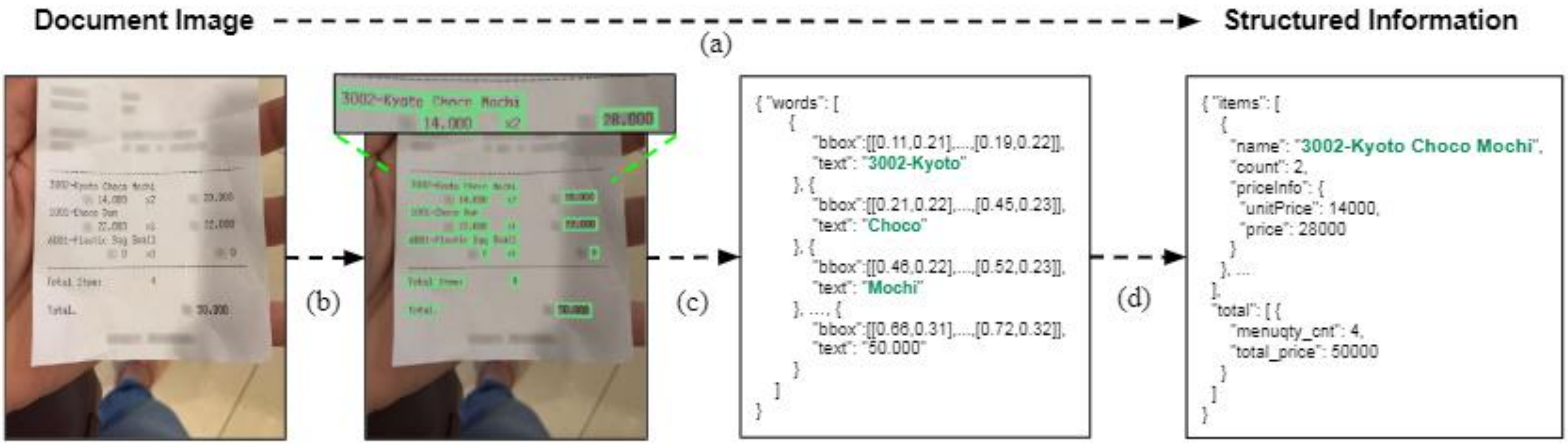


SOUP		0.31
PEAR QUARTRS		0.42
SOUP		0.31
CAT FOOD	*	2.15
ORANGE JUICE	*	0.38
OPTIONS DRINK		2.34
WHEAT BISCUITS		0.69
TESCO CRISPS	*	0.79
INSTANT NOODLE		0.21
CAKE SLICES		0.76
T W/SOME ROLLS		
REDUCED PRICE		0.35
EGGS		0.84
KM SQUARE MED		0.83
SUB-TOTAL		35.42

MULTIBUY SAVINGS

FROZEN CHICKEN 2 FOR £3	-0.48
CHICKEN TONIGHT BOGOF	-1.16
CANTALOUPE MELON BOGOF	-1.47
TOTAL SAVINGS	-3.11

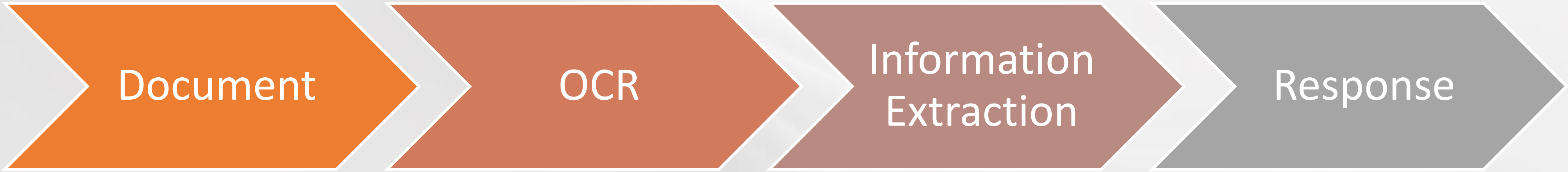
Document Processing Pipeline



```
{ "words": [
    {
        "bbox":[[0.11,0.21],...,[0.19,0.22]],
        "text": "3002-Kyoto"
    }, {
        "bbox":[[0.21,0.22],...,[0.45,0.23]],
        "text": "Choco"
    }, {
        "bbox":[[0.46,0.22],...,[0.52,0.23]],
        "text": "Mochi"
    }, ...., {
        "bbox":[[0.66,0.31],...,[0.72,0.32]],
        "text": "50.000"
    }
  ]
}
```

```
{ "items": [
    {
     "name": "3002-Kyoto Choco Mochi",
     "count": 2,
     "priceInfo": {
      "unitPrice": 14000,
      "price": 28000
     }
    }, ....
  ],
  "total": [ {
     "menuqty_cnt": 4,
     "total_price": 50000
    }
  ]
}
```

Traditional Document Processing Pipeline



Document

OCR

Information Extraction

Response

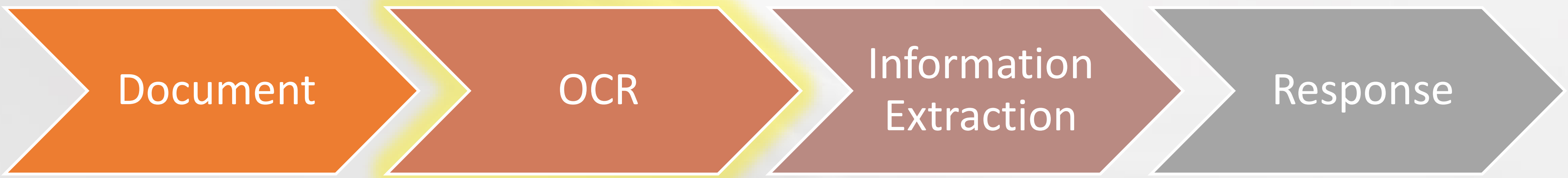
- Image extraction
- PDF Text extraction

- Image Pre-process
- Text Post-process

- NER Models
- Results Post Process

Format (JSON)

Traditional Document Processing Pipeline



Document

OCR

Information Extraction

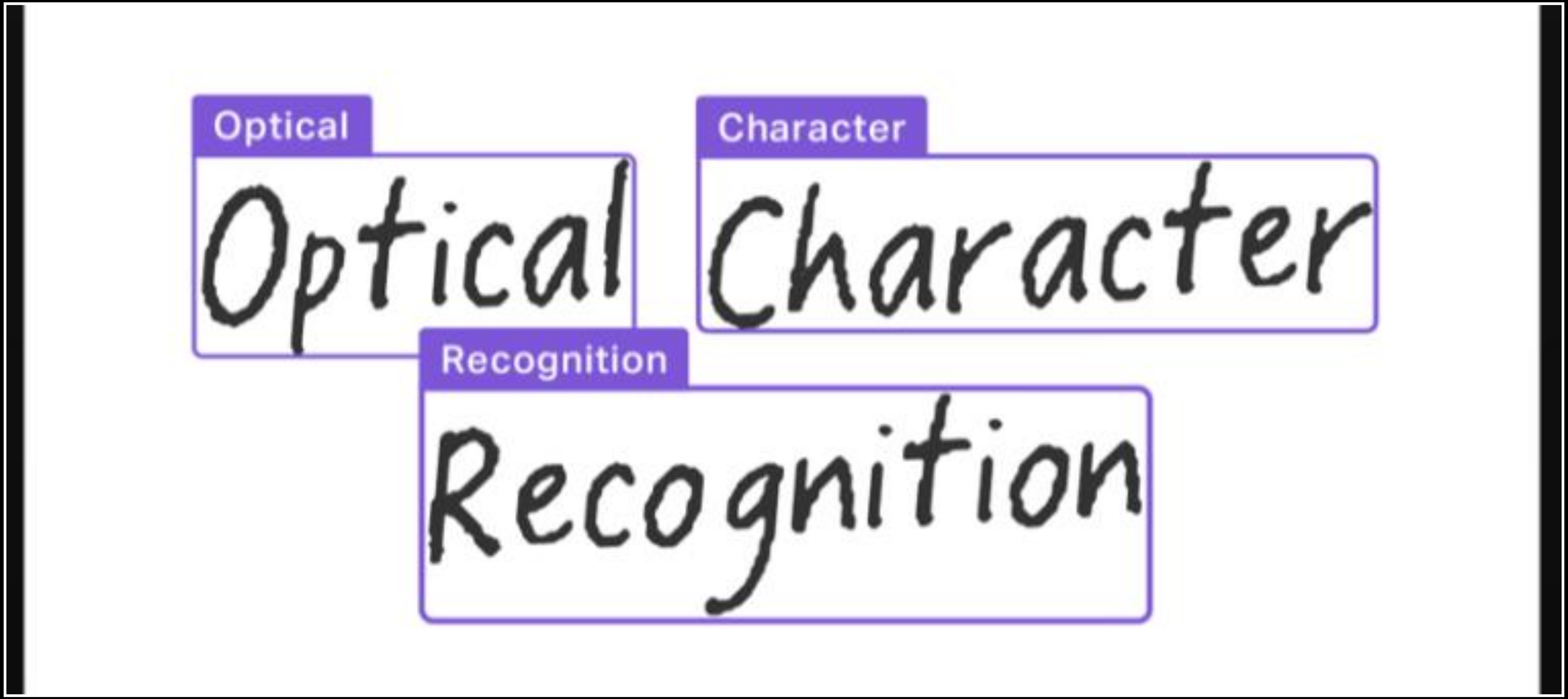
Response

- Image extraction
- PDF Text extraction

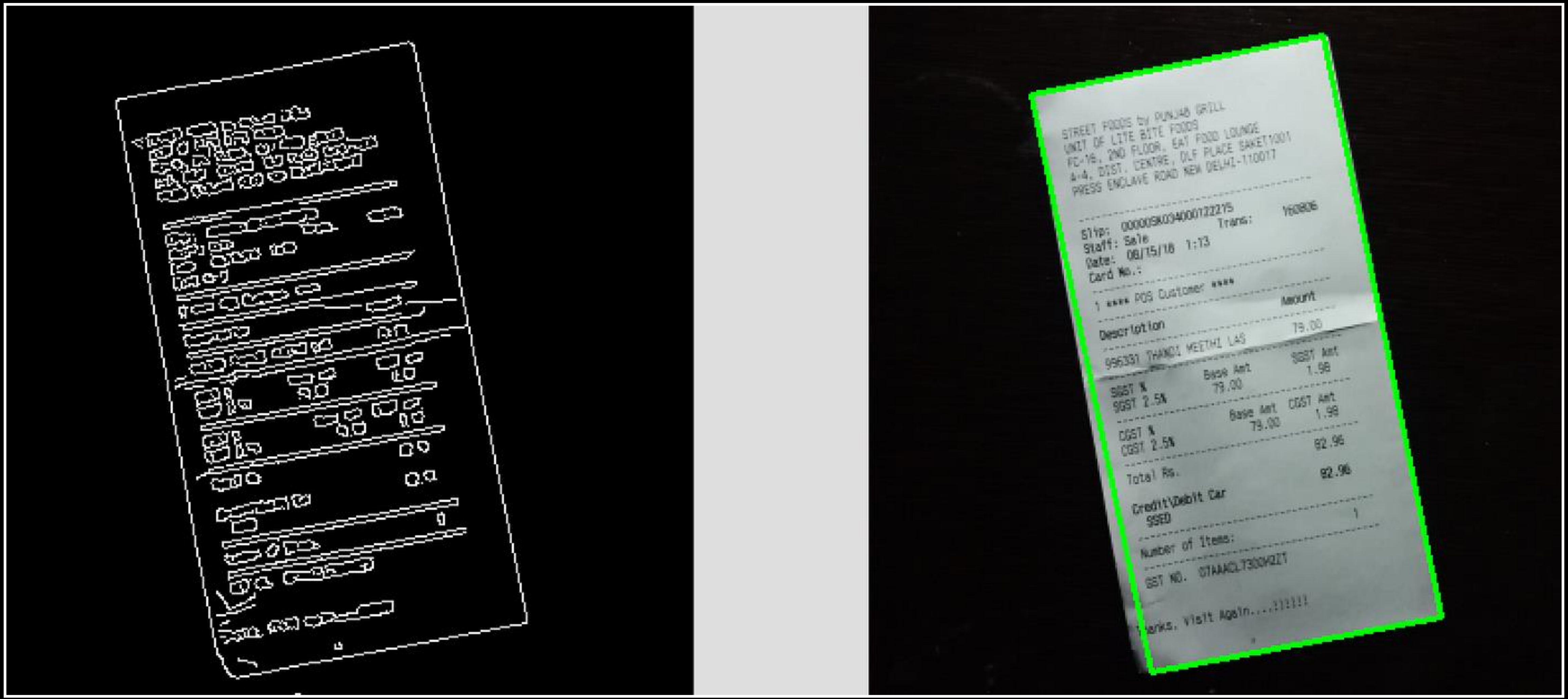
- Image Pre-process
- Text Post-process

- NER Models
- Results Post Process

Format (JSON)


Optical
Optical
Character
Character
Recognition
Recognition

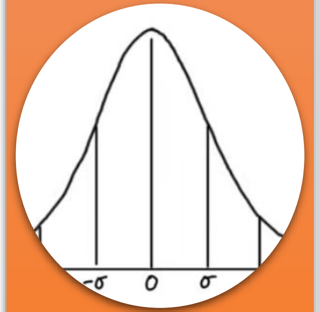
OCR



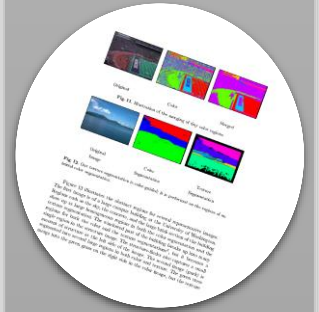
```
STREET FOODS by PUNJAB GRILL
UNIT OF LITE BITE FOODS
FC-16, 2ND FLOOR, EAT FOOD LOUNGE
A-4, DIST. CENTRE, DLF PLACE SAKET1001
PRESS ENCLAVE ROAD NEW DELHI-110017
------------------------------------
Slip:  0000SK034000122275
                          Trans:      Y60806
Staff: Sale
Date:  08/15/18  1:13
Card No.:
------------------------------------
1 **** POS Customer ****
                                  Amount
------------------------------------
Description                        79.00
------------------------------------
996331 THANDI MEETHI LAS
                                SGST Amt
                 Base Amt          1.98
SGST %             79.00
SGST 2.5%
------------------------------------
                    Base Amt  CGST Amt
CGST %                 79.00      1.98
CGST 2.5%
------------------------------------
                                   92.96
Total Rs.                          92.96

Credit\Debit Car
  9980
------------------------------------
                                       1
Number of Items:
------------------------------------
GST NO.  07AAACL7300K2Z7

anks, Visit Again.....!!!!!!
```

Often requires preprocessing

OCR: Preprocessing Steps



Normalization



Skew Correction

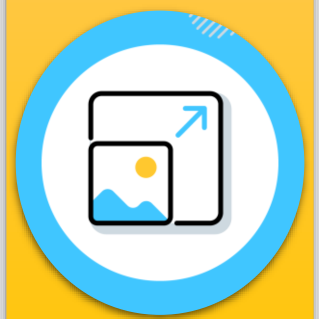
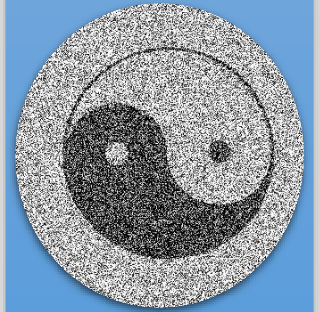
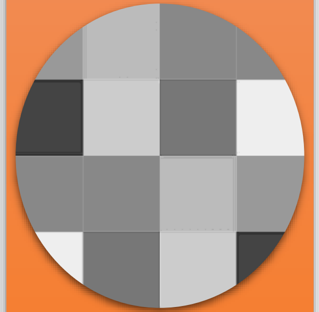


Image Scaling



Noise Removal

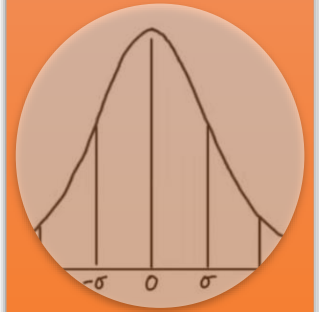
Morphological Operations: Thinning, Erosion and Dilation



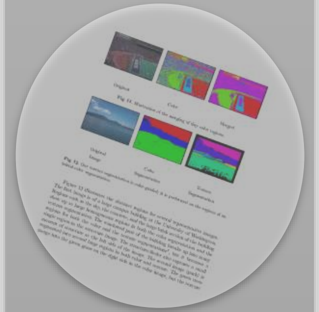
Gray scaling

Thresholding / Binarization

OCR: Preprocessing Steps

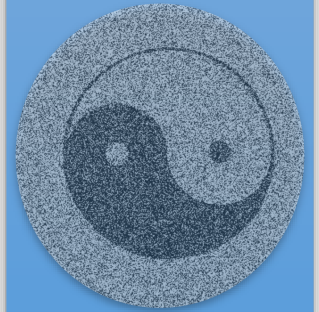


Normalization



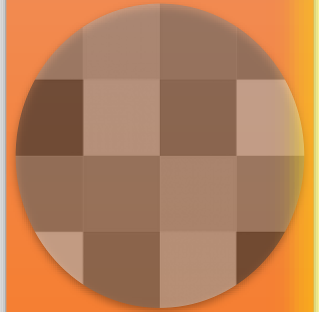
Skew Correction

Image Scaling



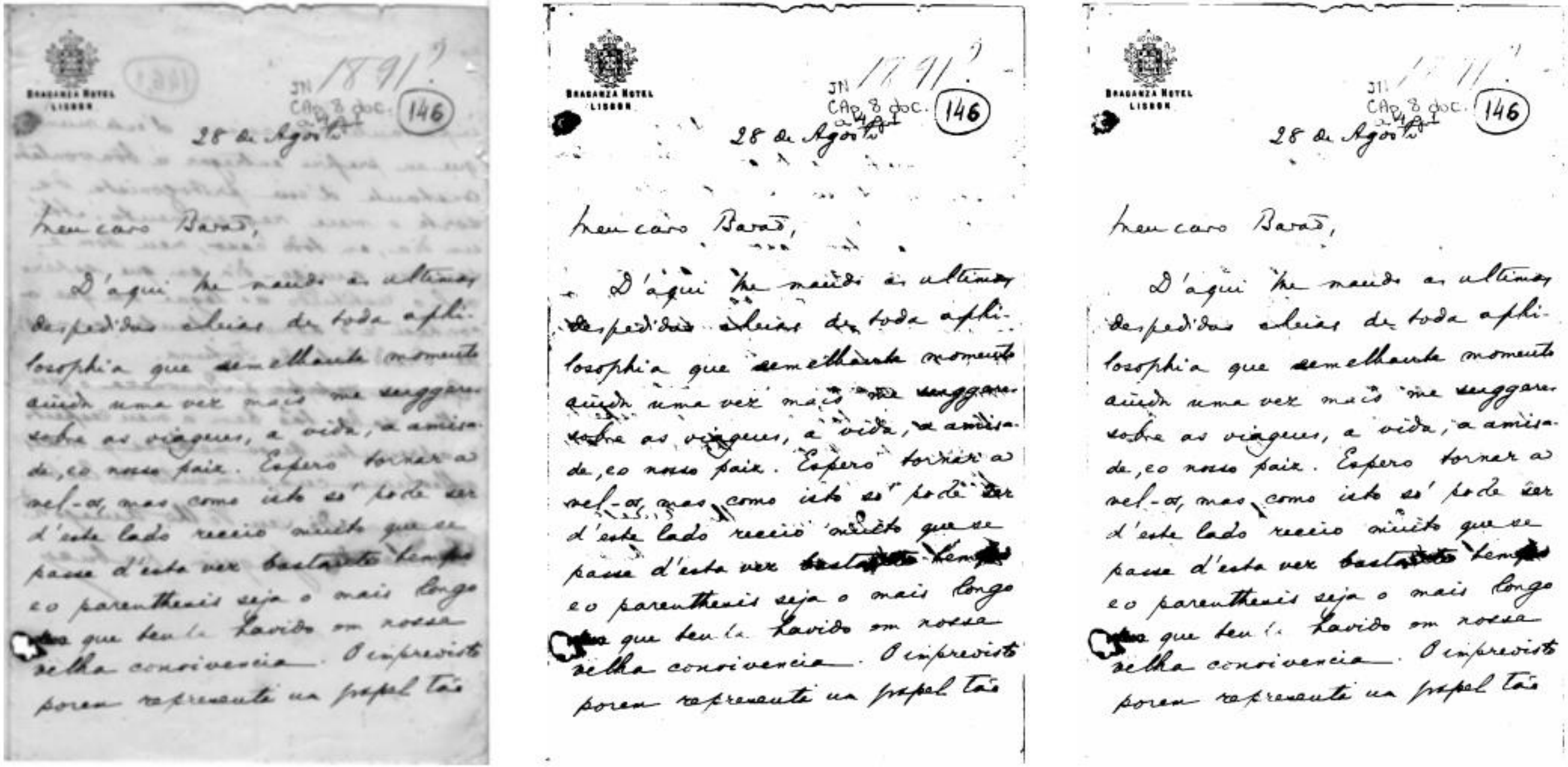
Noise Removal

Morphological Operations: Thinning, Erosion and Dilation



Gray scaling

Thresholding / Binarization



BRAGANZA HOTEL
LISBON

JN 1891?
CAp. 8 doc.
146

28 de Agosto

Meu caro Barão,

D'aqui lhe mando as ultimas despedidas cheias de toda a philosophia que semelhante momento ainda uma vez mais me suggere sobre as viagens, a vida, a amisade, e o nosso paiz. Espero tornar a vel-os, mas como isto só pode ser d'este lado receio muito que se passe d'esta vez ~~bastante~~ tempo e o parenthesis seja o mais longo que tenha havido em nossa velha convivencia. O imprevisto porem representa um papel tão

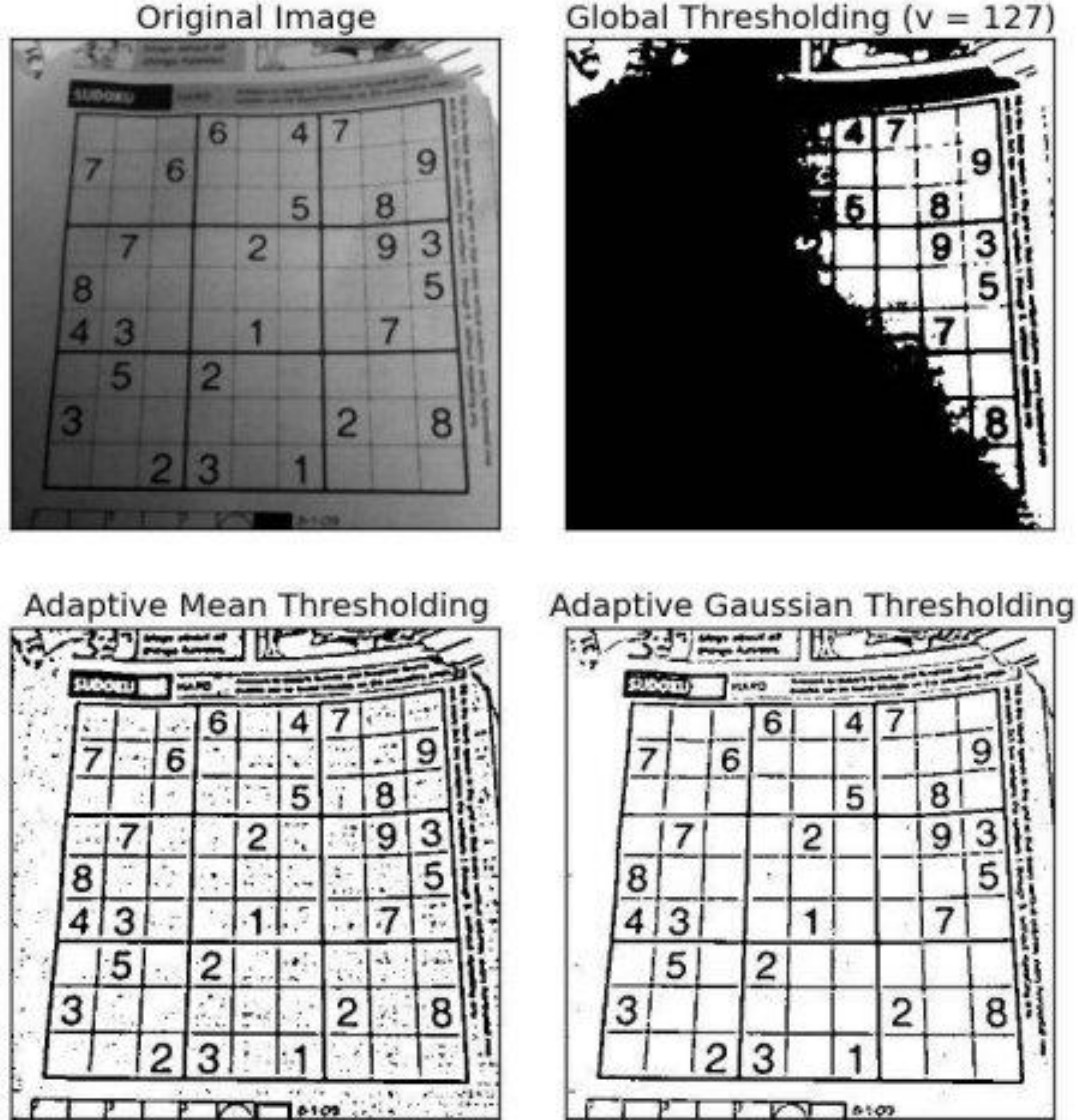
Case Study: Thresholding

- Global:
  Two-peaks, Otsu, Entropy, …
- Global Multi-Threshold
  Liao, Kapur, …
- Local:
  Local-Mean, Gaussian, Sauvola, Wolf, …

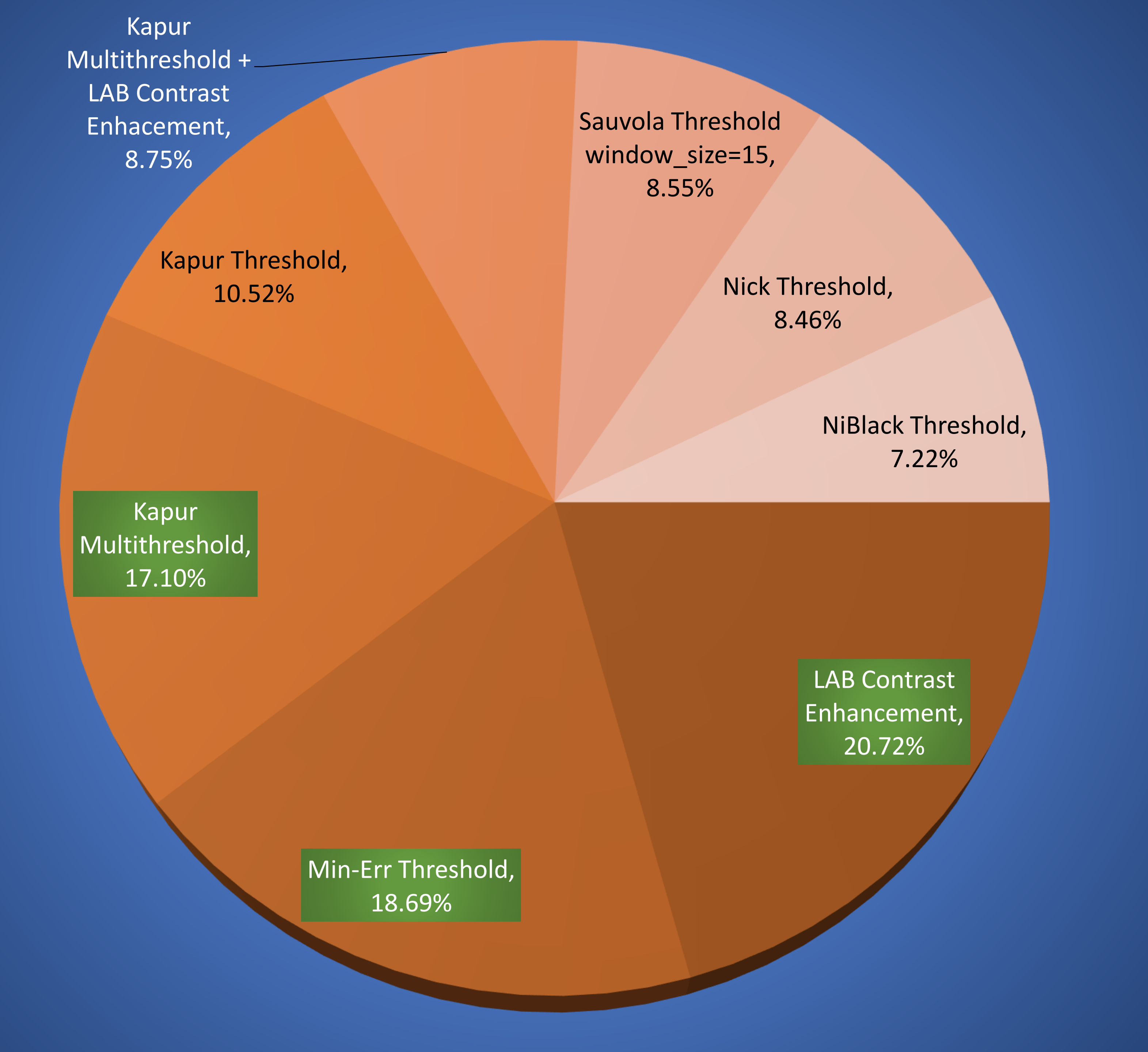
OCR Engine: **Tesseract**

Metric: **Character Error Rate (CER)**

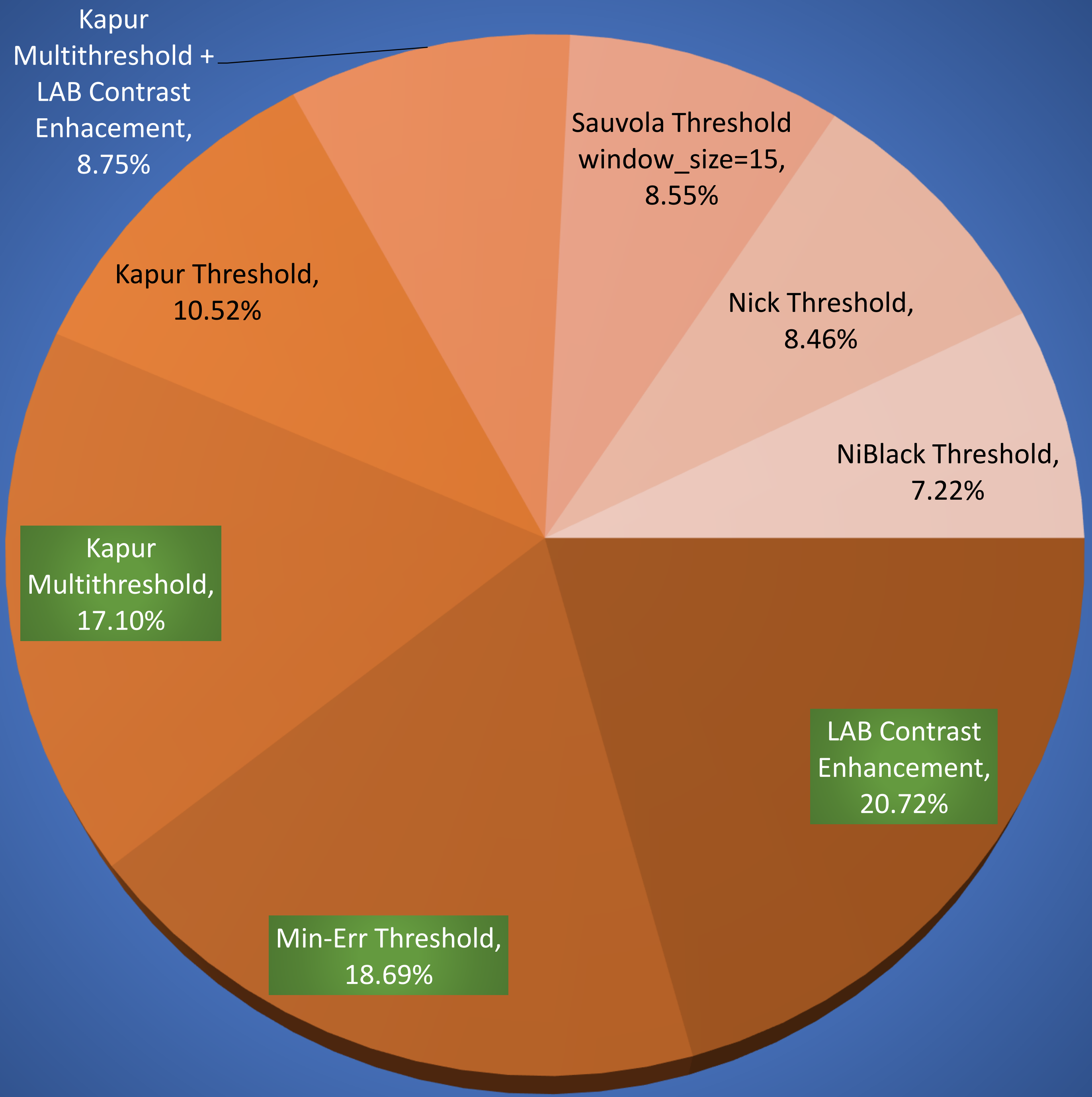
$$\boldsymbol{CER = \frac{i_c + s_c + d_c}{n_c}}$$


Original Image
Global Thresholding (v = 127)
Adaptive Mean Thresholding
Adaptive Gaussian Thresholding

Best CER Results Per Thresholding Method



How to Choose?



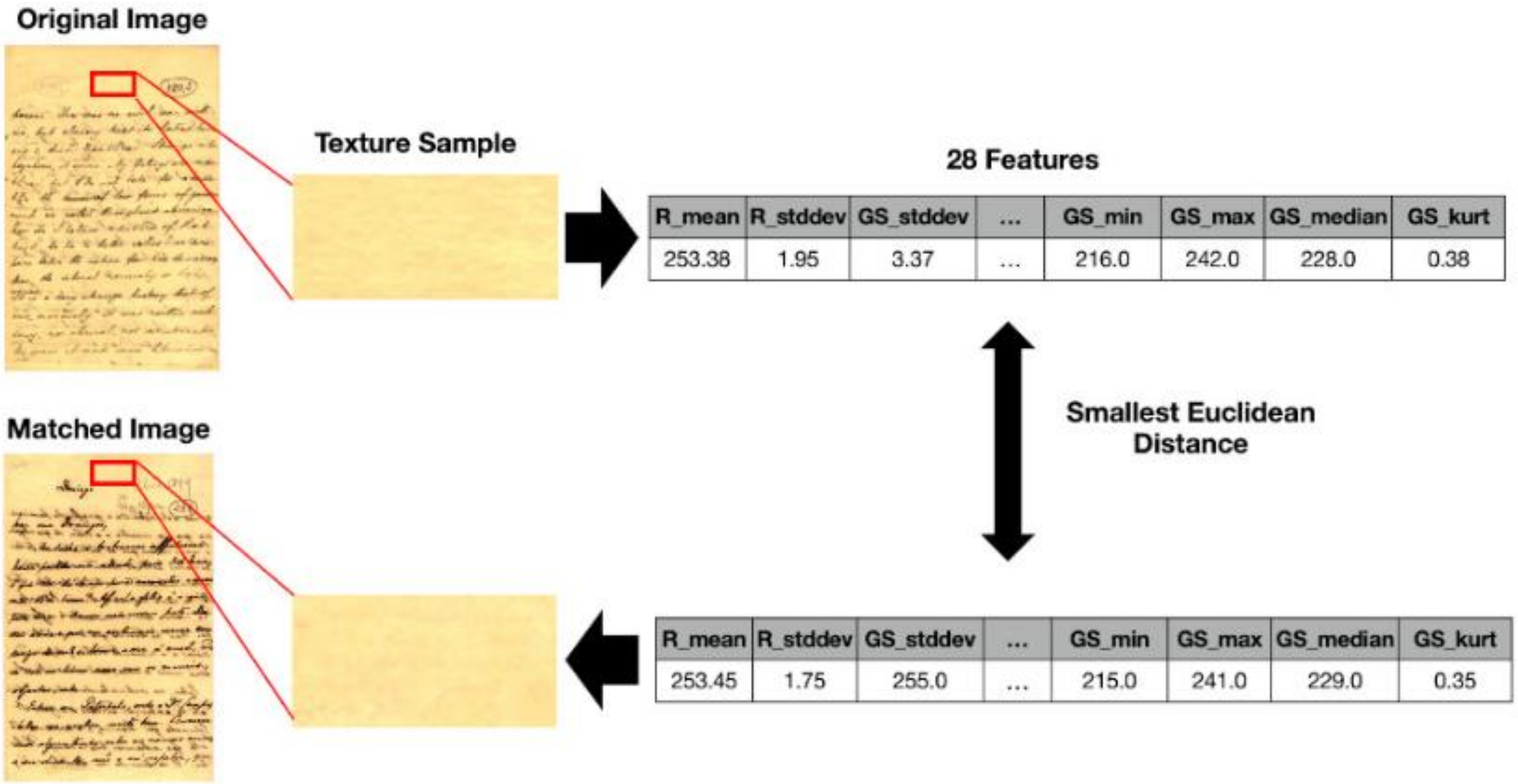
How to Choose?

“Most of the submitted algorithms employed machine learning techniques and performed best on the most complex images.

Traditional algorithms provided **very good results** at a **fraction of the time**”

ICDAR 2019 Time-Quality Binarization Competition | IEEE Conference Publication | IEEE Xplore

How to Choose?



Original Image

Texture Sample

28 Features

R_mean	R_stddev	GS_stddev	...	GS_min	GS_max	GS_median	GS_kurt
253.38	1.95	3.37	...	216.0	242.0	228.0	0.38

Smallest Euclidean Distance

Matched Image

R_mean	R_stddev	GS_stddev	...	GS_min	GS_max	GS_median	GS_kurt
253.45	1.75	255.0	...	215.0	241.0	229.0	0.35

J. Imaging | Free Full-Text | Using Paper Texture for Choosing a Suitable Algorithm for Scanned Document Image Binarization (mdpi.com)

How to Choose?

Still an open question.

OCR Post-Processing

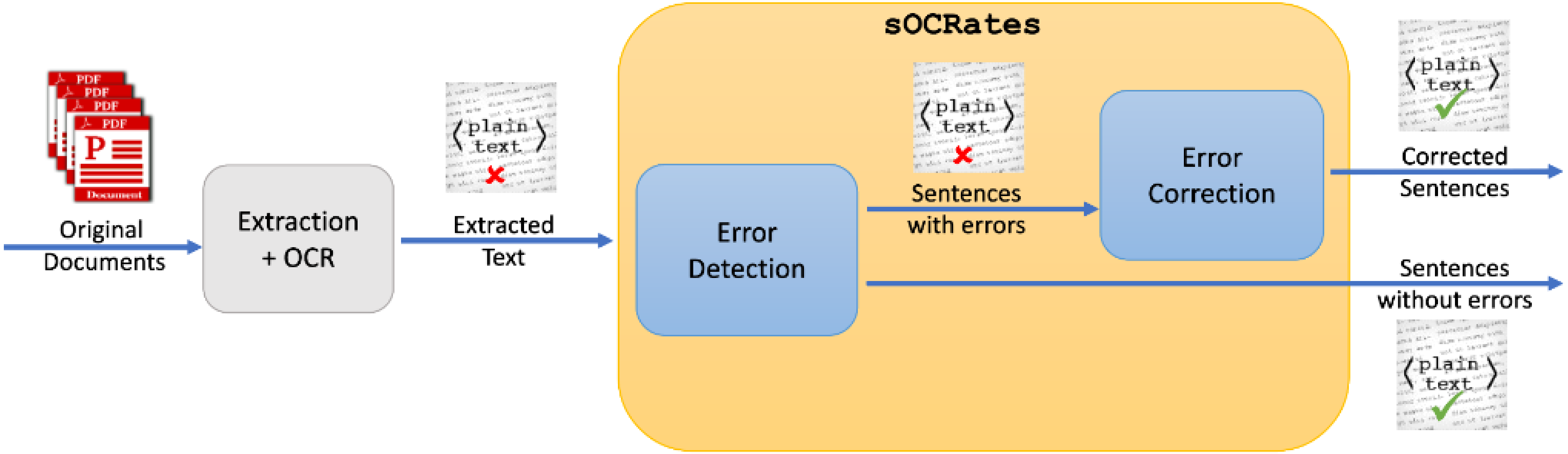


Figure 1. sOCRates and the pipeline for OCR extraction and correction.

sOCRates - a post-OCR text correction method | ScienceGate

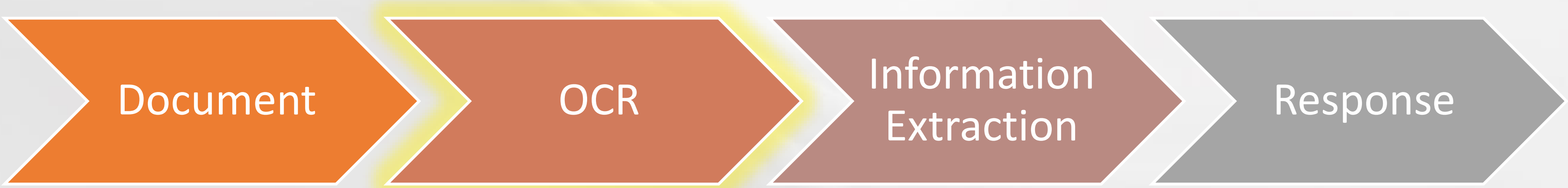
OCR: Post-processing

Error correction

- Lexical - 40% non-words
- Context-dependent (Language Models)
- Seq2Seq - Translation

Survey of Post-OCR Processing Approaches | ACM Computing Surveys

Traditional Document Processing Pipeline



Document

- Image extraction
- PDF Text extraction

OCR

- Image Pre-process
- Text Post-process

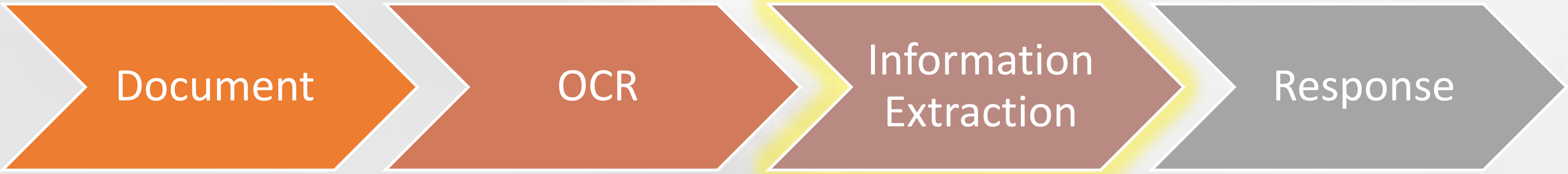
Information Extraction

- NER Models
- Results Post Process

Response

Format (JSON)

Traditional Document Processing Pipeline



Document

OCR

Information Extraction

Response

- Image extraction
- PDF Text extraction

- Image Pre-process
- Text Post-process

- NER Models
- Results Post Process

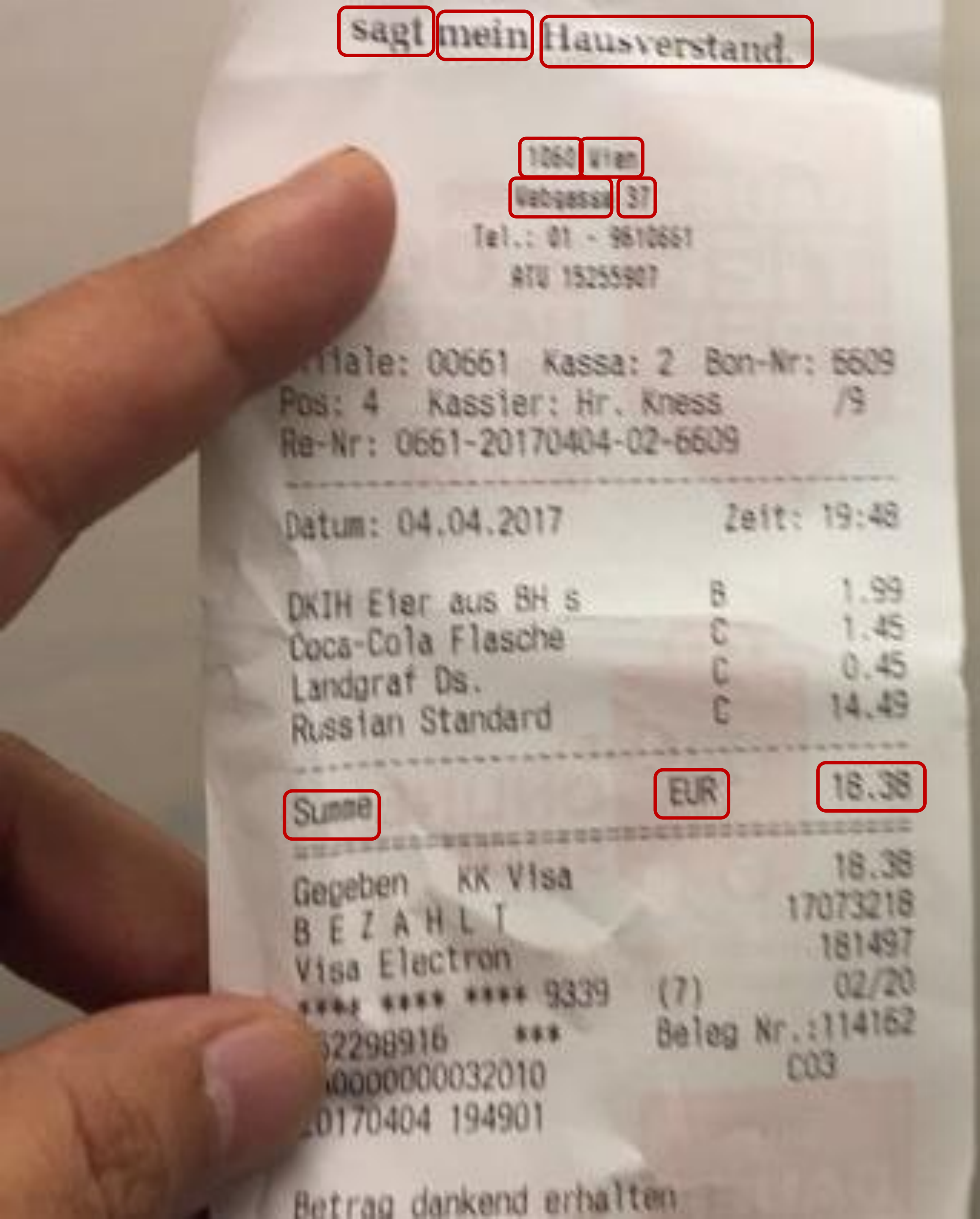
Format (JSON)

Named Entity Recognition (NER)

Named Entity Recognition (NER)

- Input: List of tokens (+ Position/Bounding Boxes)
- Output: List of Labels:
  - Invoice Date
  - Brut amount
  - Net amount
  - Tax percentage
  - Tax value
  - Recipient
  - Sender

  …



sagt mein Hausverstand.

1060 Wien
Webgasse 37
Tel.: 01 - 9610661
ATU 15255907

iliale: 00661 Kassa: 2 Bon-Nr: 6609
Pos: 4 Kassier: Hr. Kness /9
Re-Nr: 0661-20170404-02-6609

Datum: 04.04.2017 Zeit: 19:48

DKIH Eier aus BH s	B	1.99
Coca-Cola Flasche	C	1.45
Landgraf Ds.	C	0.45
Russian Standard	C	14.49
Summe	EUR	18.38

Gegeben KK Visa 18.38
B E Z A H L T 17073218
Visa Electron 181497
**** **** **** 9339 (7) 02/20
2298916 *** Beleg Nr.:114162
0000000032010 C03
0170404 194901

Betrag dankend erhalten

Named Entity Recognition (NER) - BIO

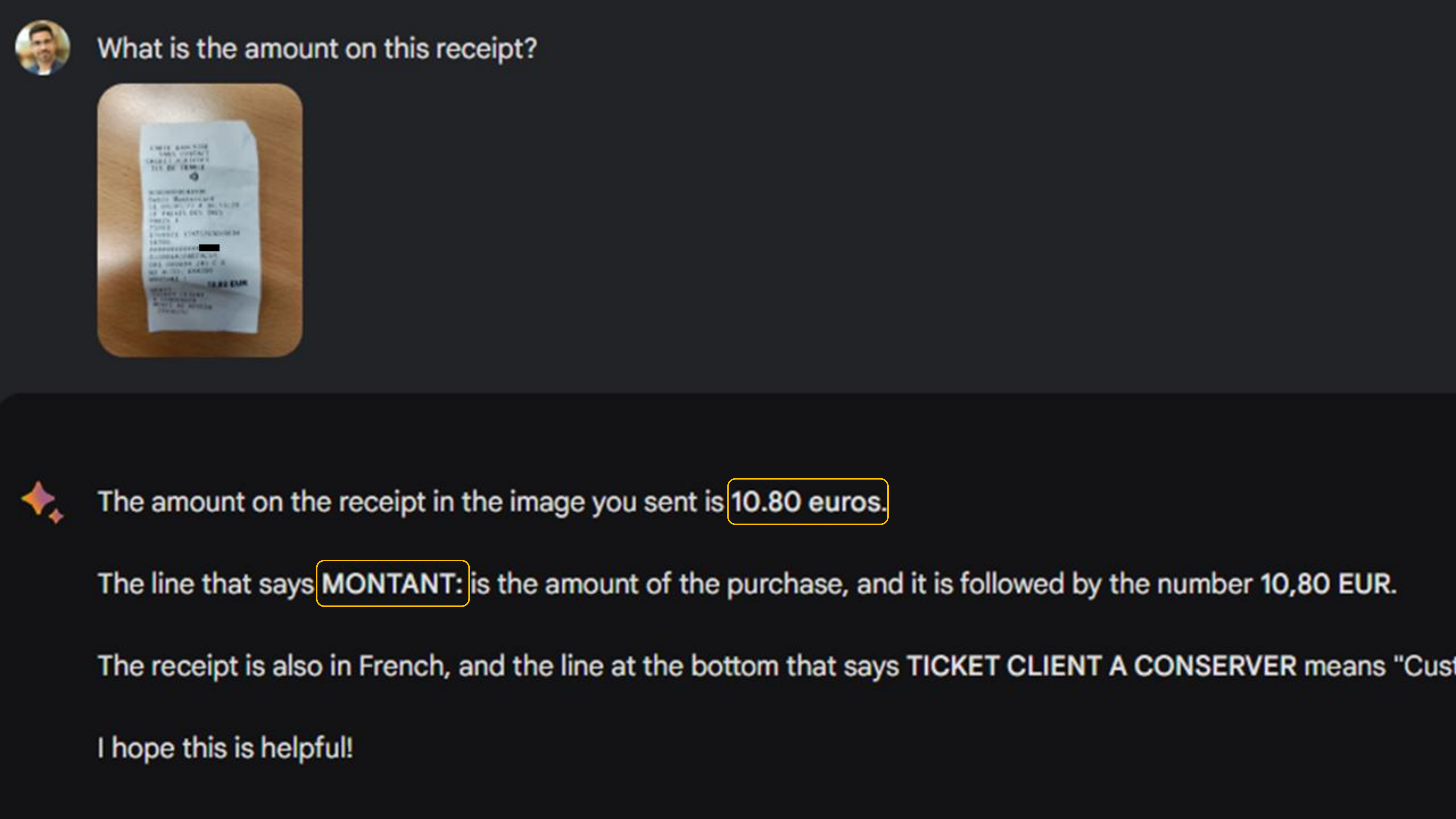
BIO Format:

BILLA	B-Issuer
Sagt	O
Mein	O
Hausverstand	O
1060	B-Zip
Wien	B-City
Webgasse	B-Address
37	I-Address

NER Post-Processing

Filter and join predicted labels

But LLMs!



What is the amount on this receipt?

The amount on the receipt in the image you sent is **10.80 euros.**

The line that says **MONTANT:** is the amount of the purchase, and it is followed by the number **10,80 EUR**.

The receipt is also in French, and the line at the bottom that says **TICKET CLIENT A CONSERVER** means "Cus

I hope this is helpful!

Large Language Models

Pros:

- Out of the box
- Can perform OCR
- Can analyze PDFs
- Multilingual support

Cons:

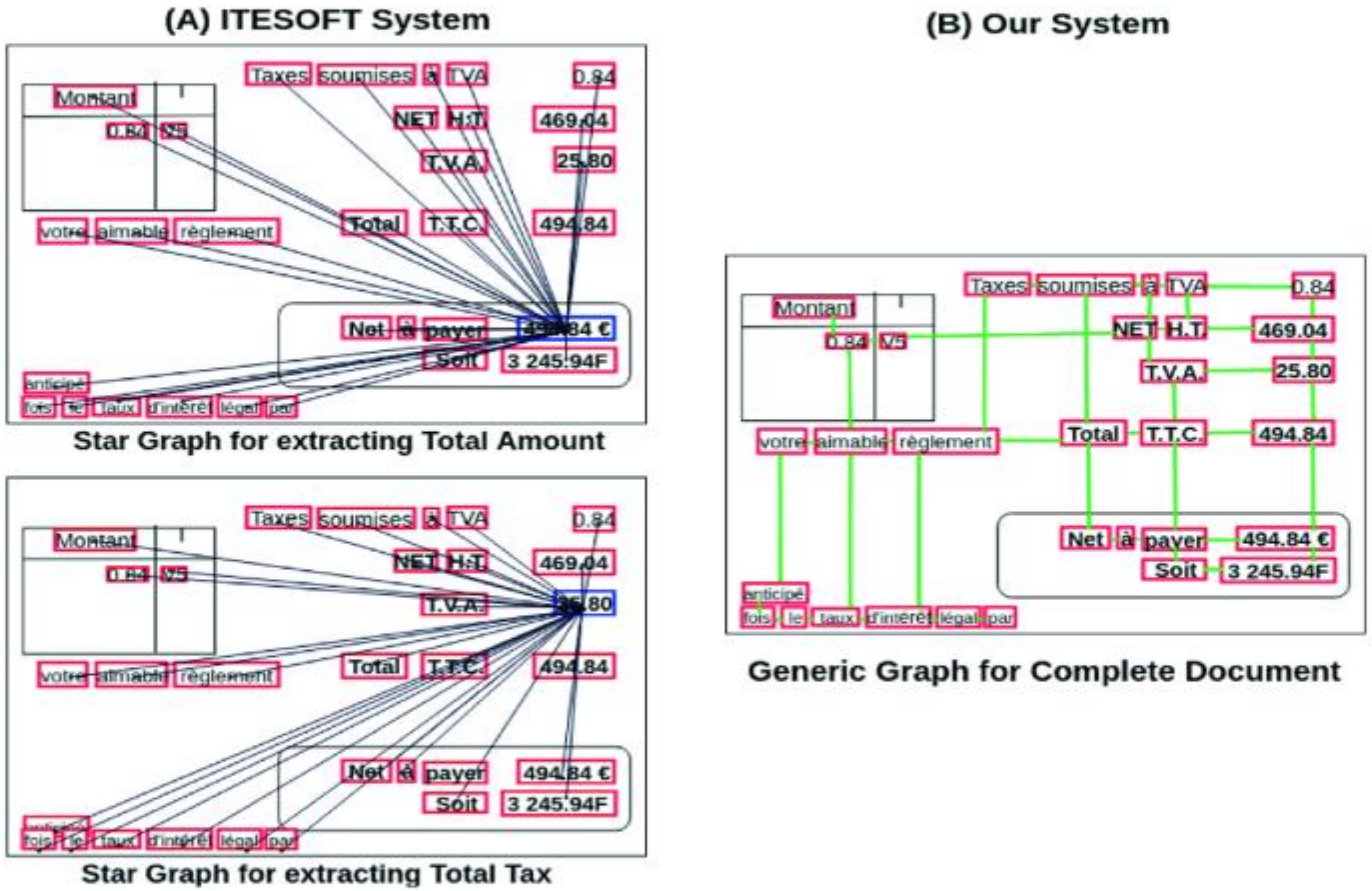
- Prompt-dependent
- Requires examples for more than basics
- Security risk:
  - Attacks
  - Prompt injections
- No on-premises option
- Reproducibility issues
  - Model Updates
  - Lack of quality control
- Liability issues

Trustworthy AI

- leiwand.ai - fair and trustworthy AI
- Consulting & Training

leiwand.ai

Position-based NER/IE Models: **GNNs**



(A) ITESOFT System

Star Graph for extracting Total Amount

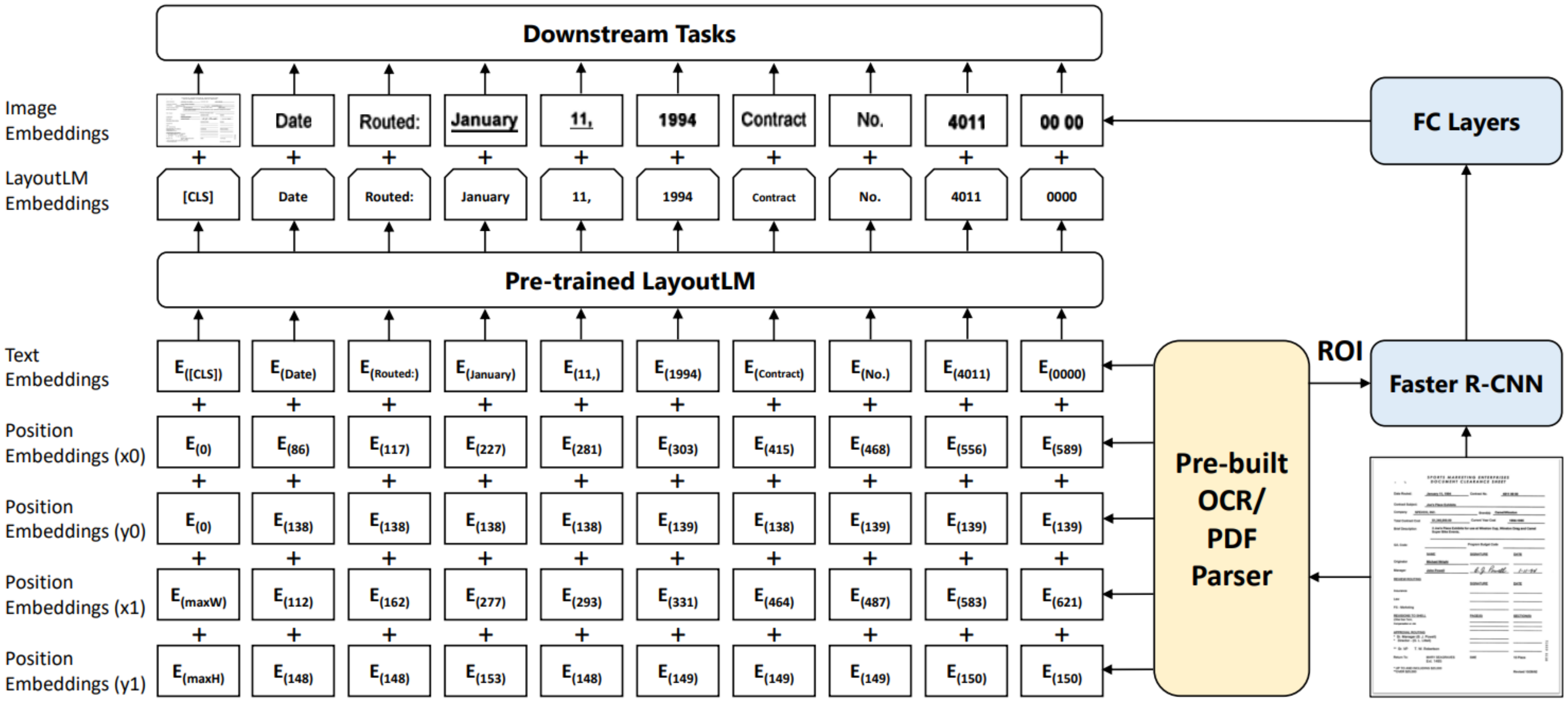
Star Graph for extracting Total Tax

(B) Our System

Generic Graph for Complete Document

An Invoice Reading System Using a Graph Convolutional Network | SpringerLink

Position-based NER/IE Models: **LayoutLM**



Downstream Tasks
Image Embeddings
Date Routed: January 11, 1994 Contract No. 4011 00 00
LayoutLM Embeddings
[CLS] Date Routed: January 11, 1994 Contract No. 4011 0000
Pre-trained LayoutLM
Text Embeddings
E([CLS]) E(Date) E(Routed:) E(January) E(11,) E(1994) E(Contract) E(No.) E(4011) E(0000)
Position Embeddings (x0)
E(0) E(86) E(117) E(227) E(281) E(303) E(415) E(468) E(556) E(589)
Position Embeddings (y0)
E(0) E(138) E(138) E(138) E(138) E(139) E(138) E(139) E(139) E(139)
Position Embeddings (x1)
E(maxW) E(112) E(162) E(277) E(293) E(331) E(464) E(487) E(583) E(621)
Position Embeddings (y1)
E(maxH) E(148) E(148) E(153) E(148) E(149) E(149) E(149) E(150) E(150)
FC Layers
ROI
Faster R-CNN
Pre-built OCR/ PDF Parser

Figure 2: An example of LayoutLM, where 2-D layout and image embeddings are integrated into the original BERT architecture. The LayoutLM embeddings and image embeddings from Faster R-CNN work together for downstream tasks.

[1912.13318] LayoutLM: Pre-training of Text and Layout for Document Image Understanding (arxiv.org)

Position-based NER/IE Models: **LayoutLMv2**

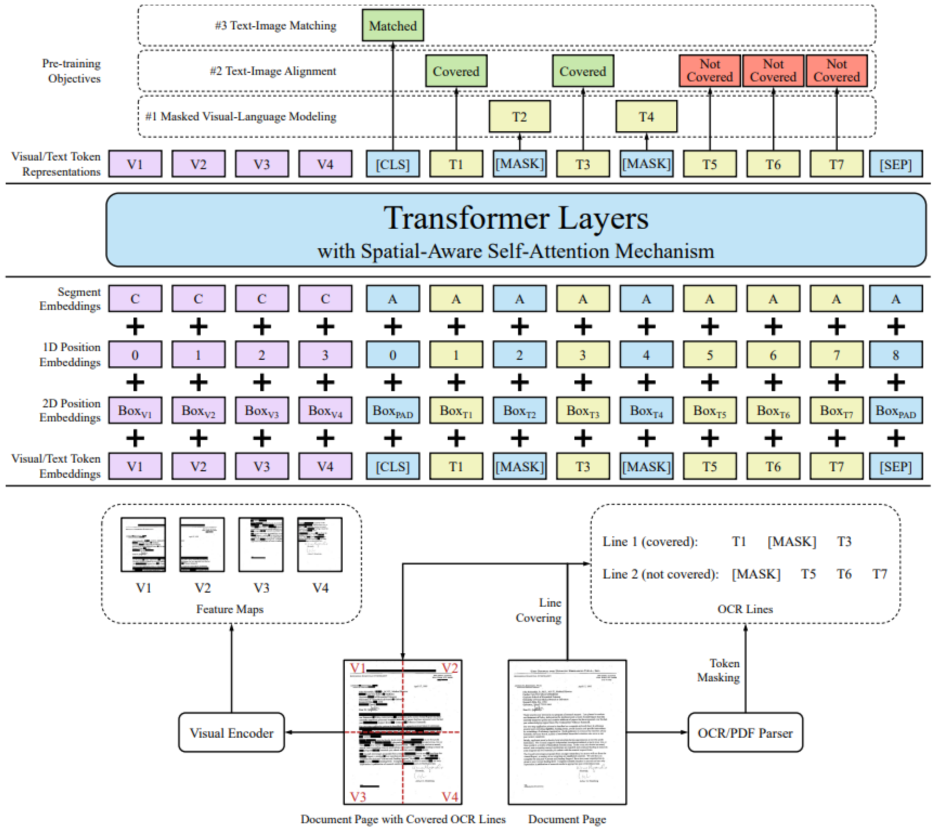
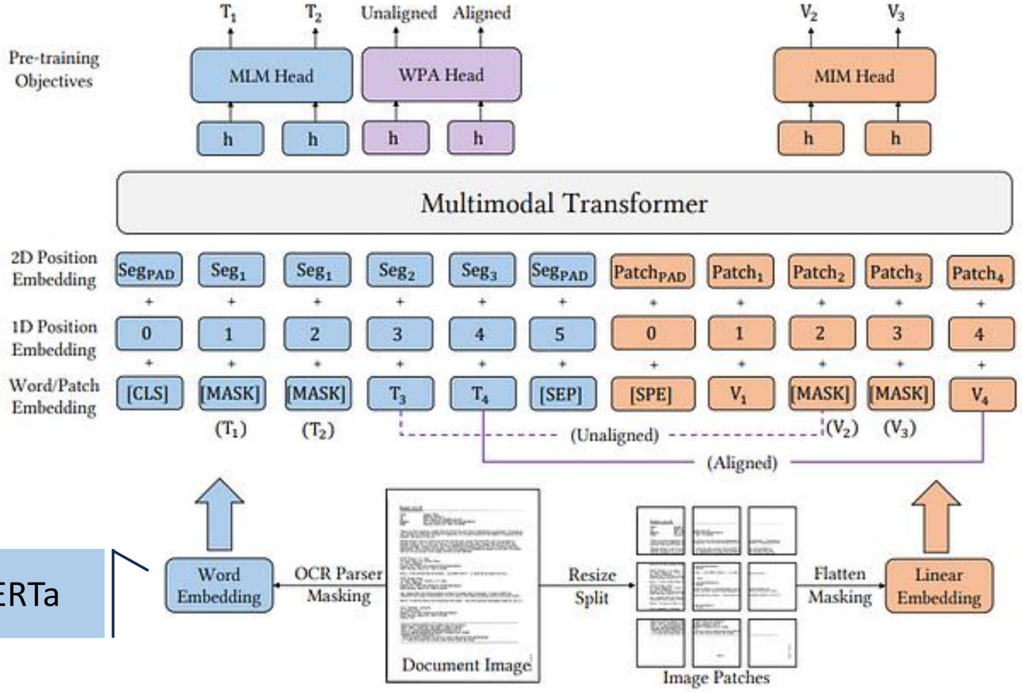

#3 Text-Image Matching
Matched
Pre-training Objectives
#2 Text-Image Alignment
Covered
Covered
Not Covered
Not Covered
Not Covered
#1 Masked Visual-Language Modeling
T2
T4
Visual/Text Token Representations
V1 V2 V3 V4 [CLS] T1 [MASK] T3 [MASK] T5 T6 T7 [SEP]
Transformer Layers
with Spatial-Aware Self-Attention Mechanism
Segment Embeddings
C C C C A A A A A A A A A
1D Position Embeddings
0 1 2 3 0 1 2 3 4 5 6 7 8
2D Position Embeddings
Box$_{V1}$ Box$_{V2}$ Box$_{V3}$ Box$_{V4}$ Box$_{PAD}$ Box$_{T1}$ Box$_{T2}$ Box$_{T3}$ Box$_{T4}$ Box$_{T5}$ Box$_{T6}$ Box$_{T7}$ Box$_{PAD}$
Visual/Text Token Embeddings
V1 V2 V3 V4 [CLS] T1 [MASK] T3 [MASK] T5 T6 T7 [SEP]
V1 V2 V3 V4
Feature Maps
Line 1 (covered): T1 [MASK] T3
Line 2 (not covered): [MASK] T5 T6 T7
OCR Lines
Line Covering
Token Masking
Visual Encoder
OCR/PDF Parser
V1 V2 V3 V4
Document Page with Covered OCR Lines
Document Page

Figure 1: An illustration of the model architecture and pre-training strategies for LayoutLMv2

[2012.14740] LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding (arxiv.org)

Position-based NER/IE Models: **LayoutLMv3**

Text Embedding with Positional Embedding



RoBERTa

Figure 3: The architecture and pre-training objectives of LayoutLMv3. LayoutLMv3 is a pre-trained multimodal Transformer for Document AI with unified text and image masking objectives. Given an input document image and its corresponding text and layout position information, the model takes the linear projection of patches and word tokens as inputs and encodes them into contextualized vector representations. LayoutLMv3 is pre-trained with discrete token reconstructive objectives of Masked Language Modeling (MLM) and Masked Image Modeling (MIM). Additionally, LayoutLMv3 is pre-trained with a Word-Patch Alignment (WPA) objective to learn cross-modal alignment by predicting whether the corresponding image patch of a text word is masked. "Seg" denotes segment-level positions. "[CLS]", "[MASK]", "[SEP]" and "[SPE]" are special tokens.

[[2204.08387] LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking (arxiv.org)]()

Position-based NER/IE Models: **BROS***

* BERT Relying on Spatiality.

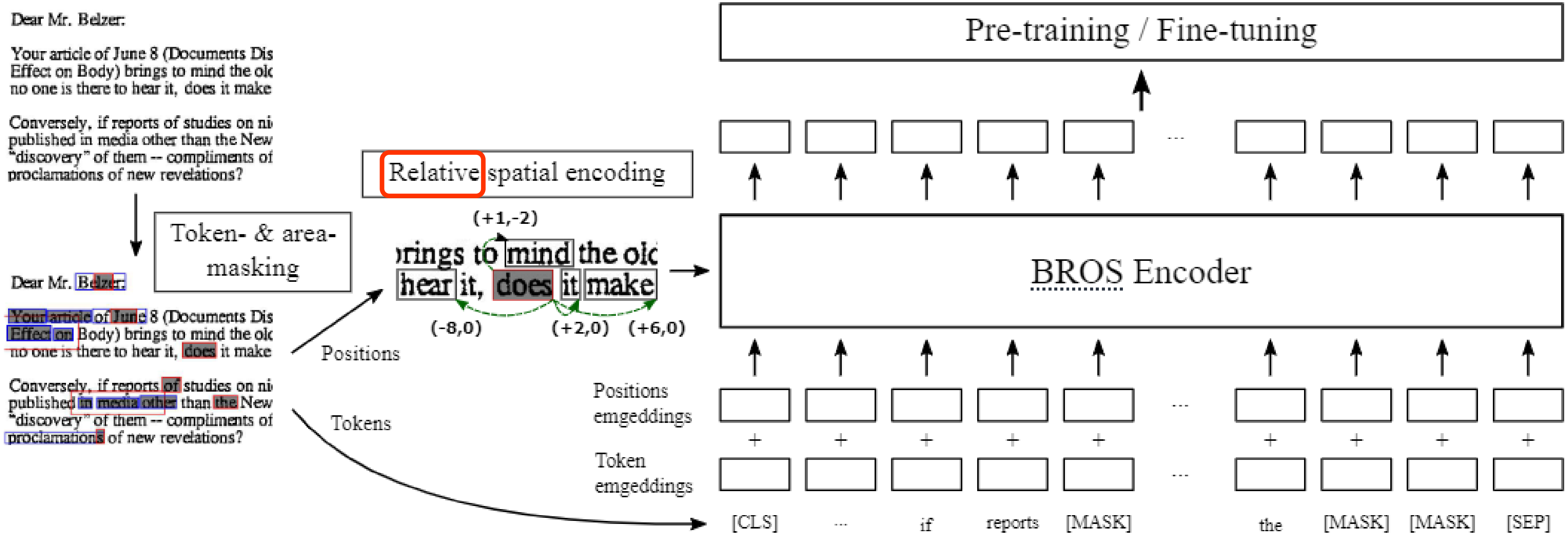

Dear Mr. Belzer:
Your article of June 8 (Documents Dis
Effect on Body) brings to mind the olc
no one is there to hear it, does it make
Conversely, if reports of studies on ni
published in media other than the New
"discovery" of them -- compliments of
proclamations of new revelations?
Token- & area-masking
Positions
Tokens
Relative spatial encoding
(+1,-2)
brings to mind the olc
hear it, does it make
(-8,0)
(+2,0)
(+6,0)
Pre-training / Fine-tuning
BROS Encoder
Positions emgeddings
Token emgeddings
[CLS]
...
if
reports
[MASK]
the
[MASK]
[MASK]
[SEP]

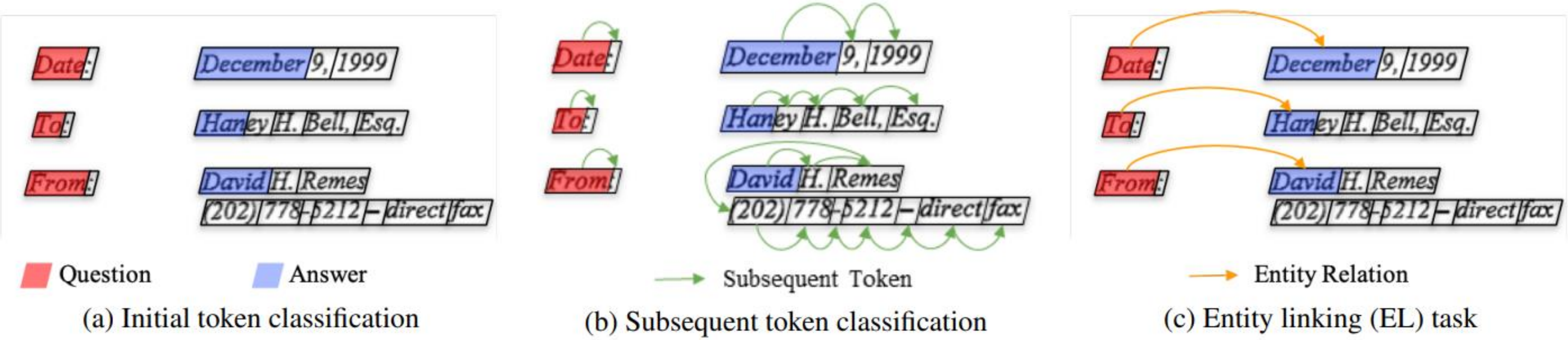
Figure 2: An overview of BROS. The tokens in the document image are masked through token- and area-masking strategy. The position difference between text blocks is encoded directly to the attention mechanism in Transformer. The output token representations are used in both pre-training and fine-tuning.

[[2108.04539] BROS: A Pre-trained Language Model Focusing on Text and Layout for Better Key Information Extraction from Documents (arxiv.org)]()

Position-based NER/IE Models: **BROS***

* BERT Relying on Spatiality.

SPADE Decoder



(a) Initial token classification

(b) Subsequent token classification

(c) Entity linking (EL) task

[[2108.04539] BROS: A Pre-trained Language Model Focusing on Text and Layout for Better Key Information Extraction from Documents (arxiv.org)]

Position-based NER/IE Models: **GeoLayoutLM**

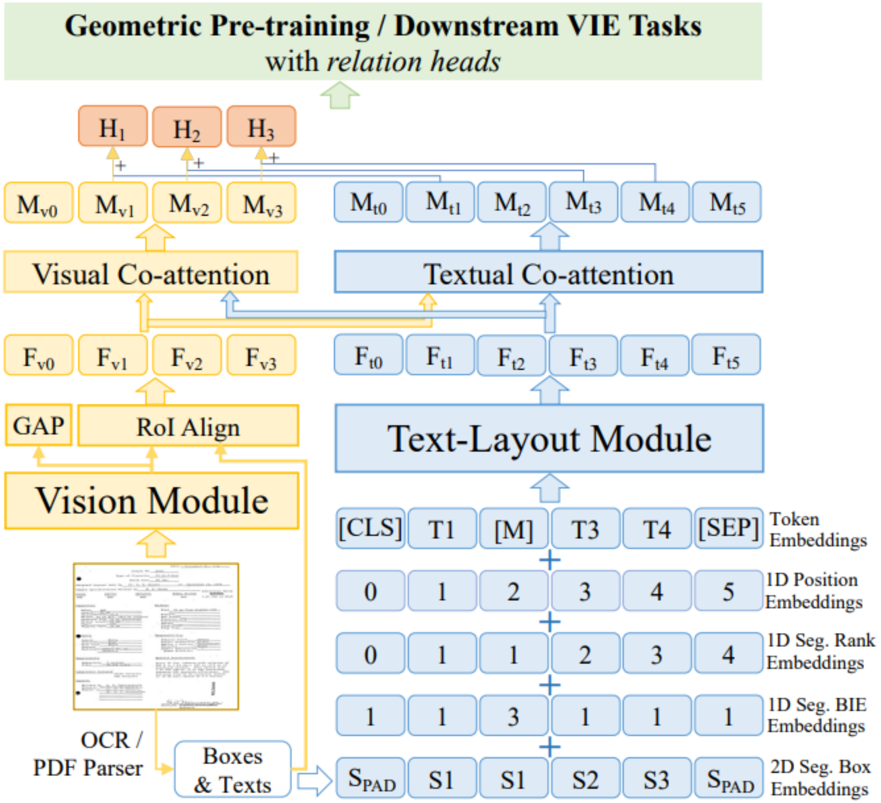
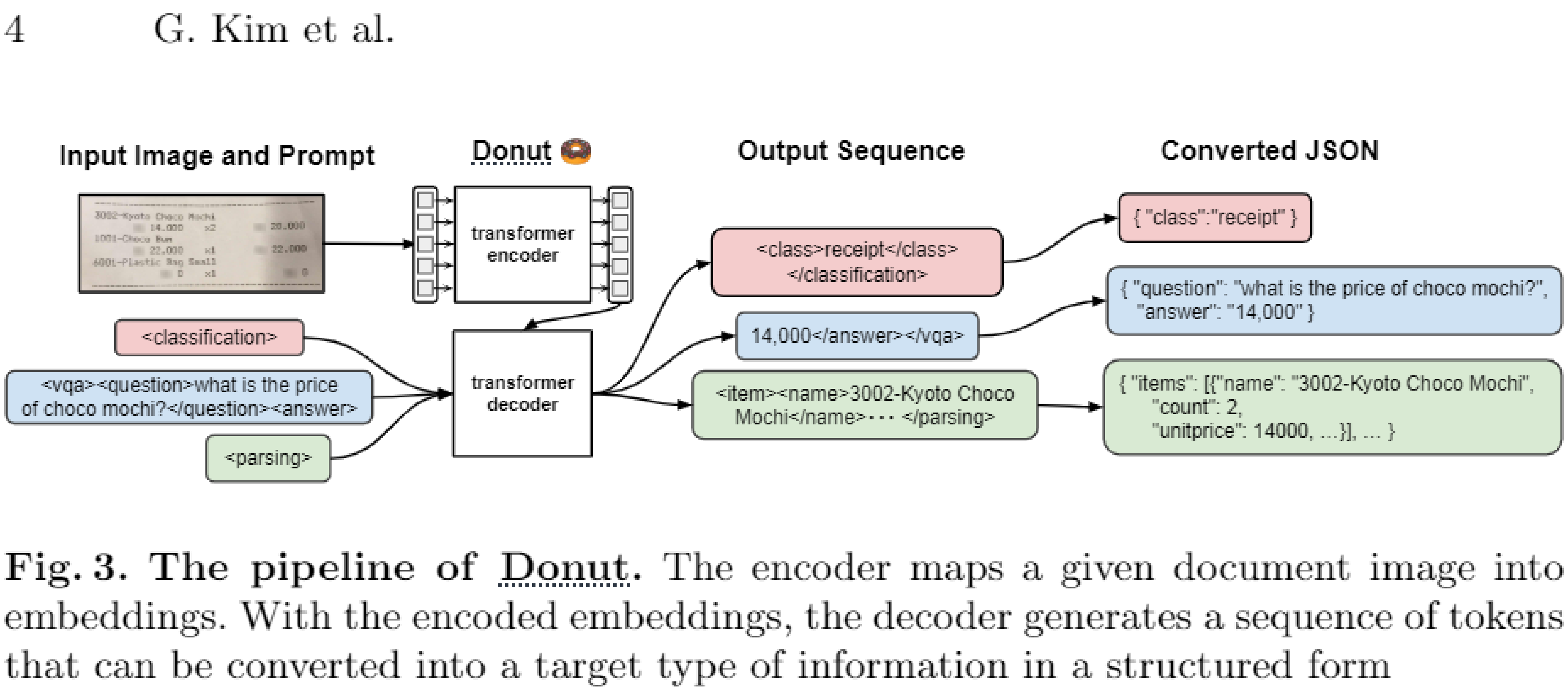


Figure 2. An overview of GeoLayoutLM.

[2304.10759] GeoLayoutLM: Geometric Pre-training for Visual Information Extraction (arxiv.org)

Position-based NER/IE Models: **Donut**

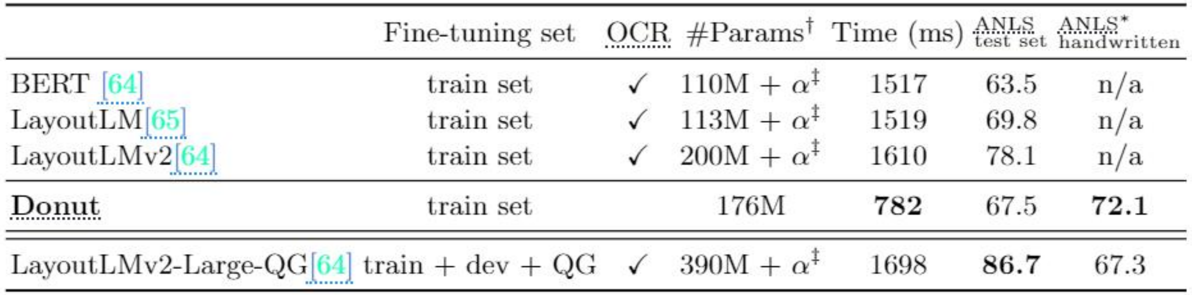


4 G. Kim et al.

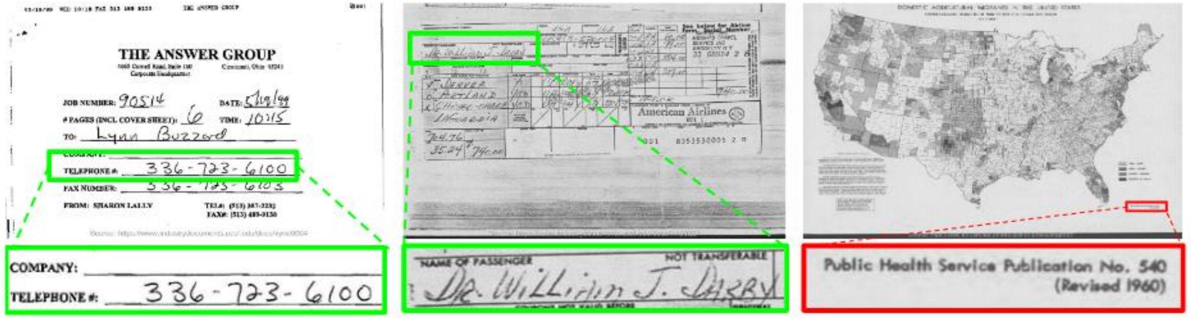
Input Image and Prompt
Donut 🍩
Output Sequence
Converted JSON
transformer encoder
transformer decoder
<classification>
<vqa><question>what is the price of choco mochi?</question><answer>
<parsing>
<class>receipt</class></classification>
14,000</answer></vqa>
<item><name>3002-Kyoto Choco Mochi</name>··· </parsing>
{ "class":"receipt" }
{ "question": "what is the price of choco mochi?", "answer": "14,000" }
{ "items": [{"name": "3002-Kyoto Choco Mochi", "count": 2, "unitprice": 14000, …}], … }

Fig. 3. The pipeline of Donut. The encoder maps a given document image into embeddings. With the encoded embeddings, the decoder generates a sequence of tokens that can be converted into a target type of information in a structured form

Position-based NER/IE Models: **Donut**



	Fine-tuning set	OCR	#Params†	Time (ms)	ANLS test set	ANLS* handwritten
BERT [64]	train set	✓	110M + $\alpha^\ddagger$	1517	63.5	n/a
LayoutLM [65]	train set	✓	113M + $\alpha^\ddagger$	1519	69.8	n/a
LayoutLMv2 [64]	train set	✓	200M + $\alpha^\ddagger$	1610	78.1	n/a
Donut	train set		176M	**782**	67.5	**72.1**
LayoutLMv2-Large-QG [64]	train + dev + QG	✓	390M + $\alpha^\ddagger$	1698	**86.7**	67.3


THE ANSWER GROUP
COMPANY:
TELEPHONE #: 336-723-6100
NAME OF PASSENGER
NOT TRANSFERABLE
Dr. William J. Darby
American Airlines
Public Health Service Publication No. 540
(Revised 1960)

Q: What is the phone number given?
Answer: 336-723-6100
Donut: 336-723-6100
LayoutLMv2-Large-QG: 336-723- 4100

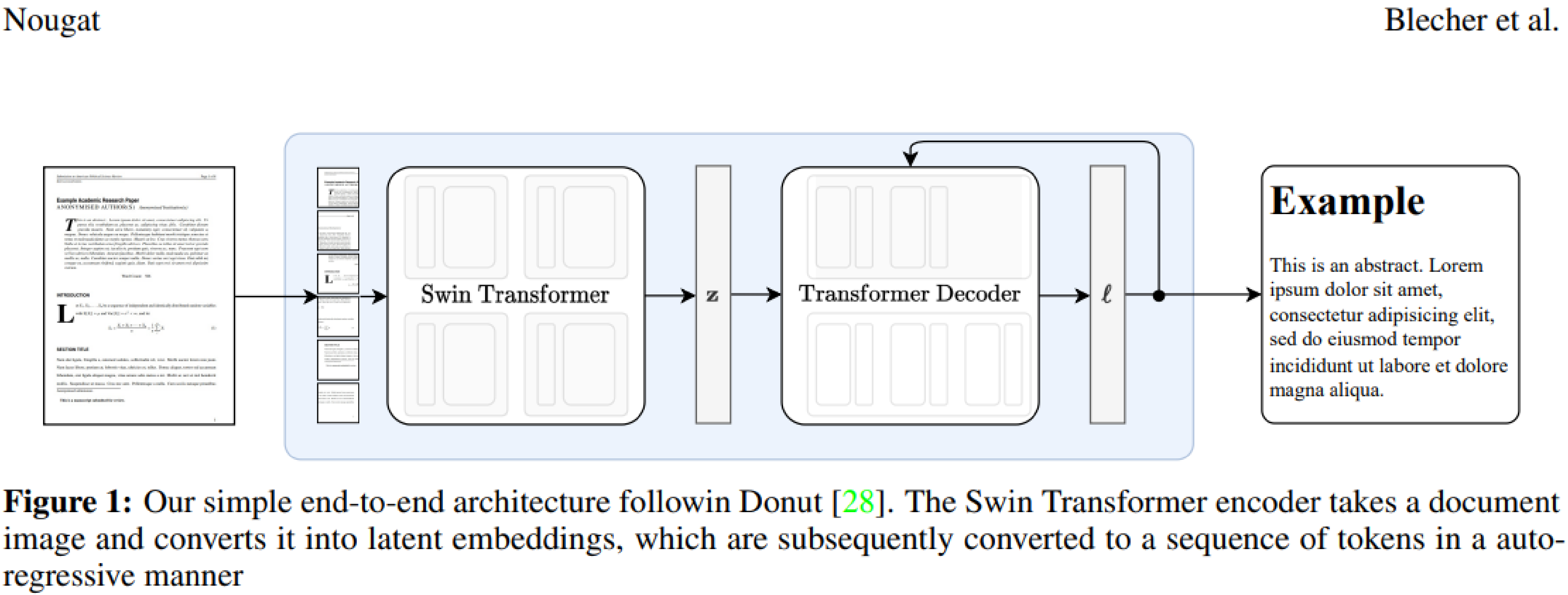
Q: What is the name of the passenger?
Answer: DR. William J. Darby
Donut: DR. William J. Darby
LayoutLMv2-Large-QG: DR. William J. Jarry

Q: What is the Publication No.?
Answer: 540
Donut: 943 (another number in the image is extracted)
LayoutLMv2-Large-QG: 540

- Pretrain + Finetune
  - SWIN Transformer

- Handwriting
- Synthetic training data

Position-based NER/IE Models: **Nougat**



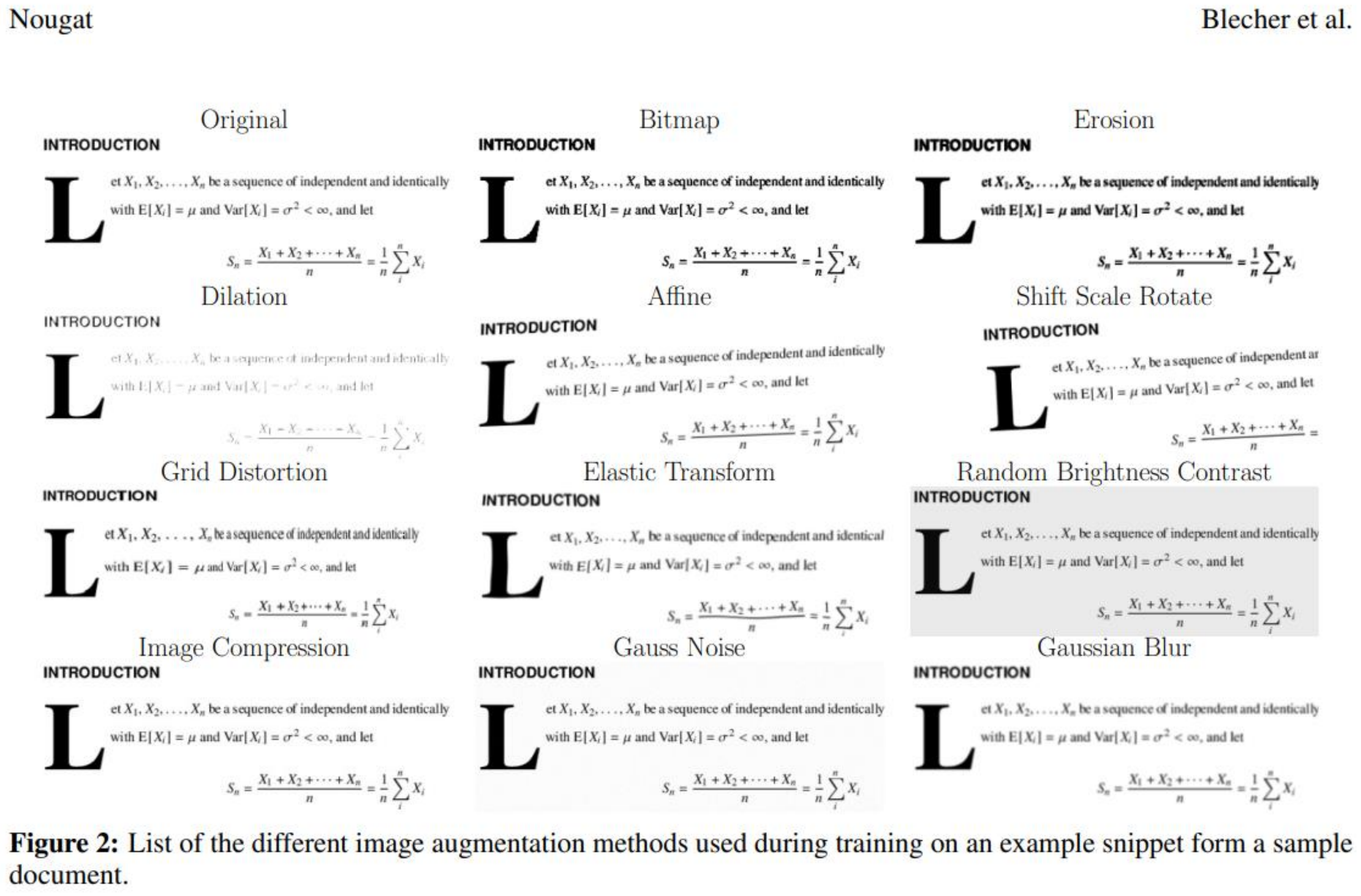
Nougat Blecher et al.

This is an abstract. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Figure 1: Our simple end-to-end architecture followin Donut [28]. The Swin Transformer encoder takes a document image and converts it into latent embeddings, which are subsequently converted to a sequence of tokens in a auto-regressive manner

[Nougat (facebookresearch.github.io)](https://facebookresearch.github.io/nougat/)

Position-based NER/IE Models: **Nougat**



Nougat Blecher et al.

Original
Bitmap
Erosion
Dilation
Affine
Shift Scale Rotate
Grid Distortion
Elastic Transform
Random Brightness Contrast
Image Compression
Gauss Noise
Gaussian Blur
INTRODUCTION
L
et $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically
with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$, and let
$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_i^n X_i$

Figure 2: List of the different image augmentation methods used during training on an example snippet form a sample document.

[Nougat (facebookresearch.github.io)](https://facebookresearch.github.io/nougat/)

Position-based NER/IE Models: **DUBLIN**

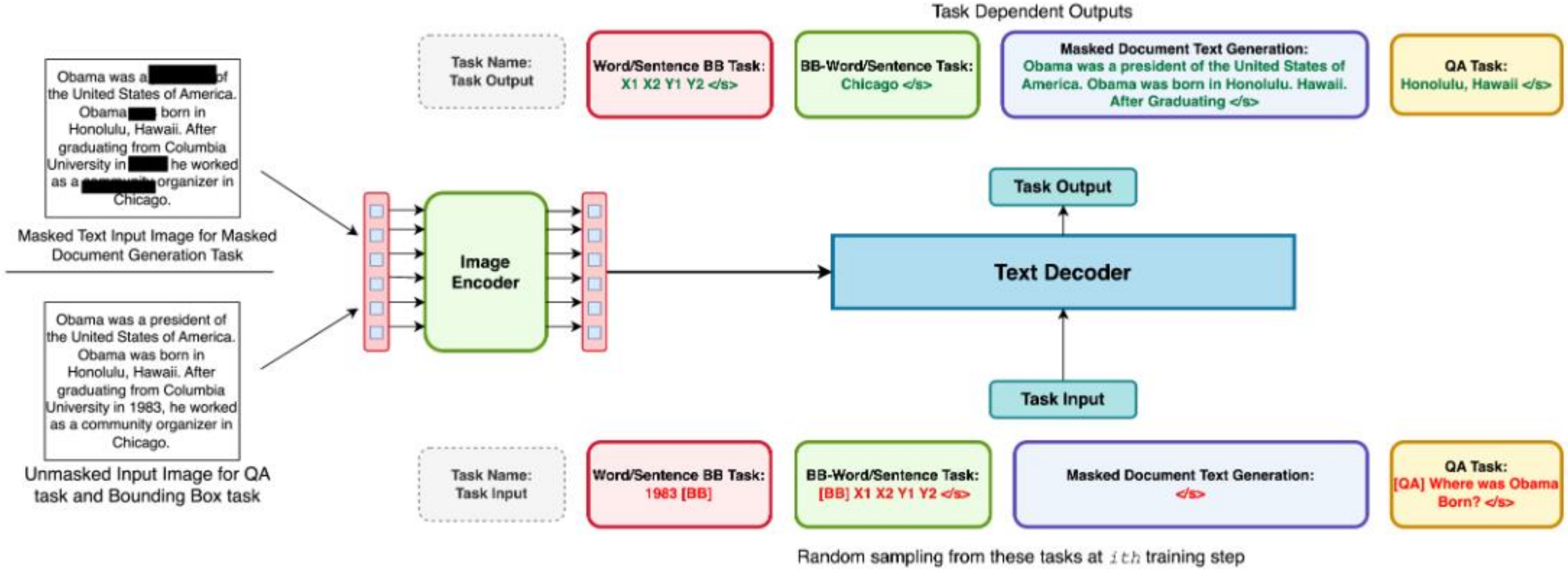
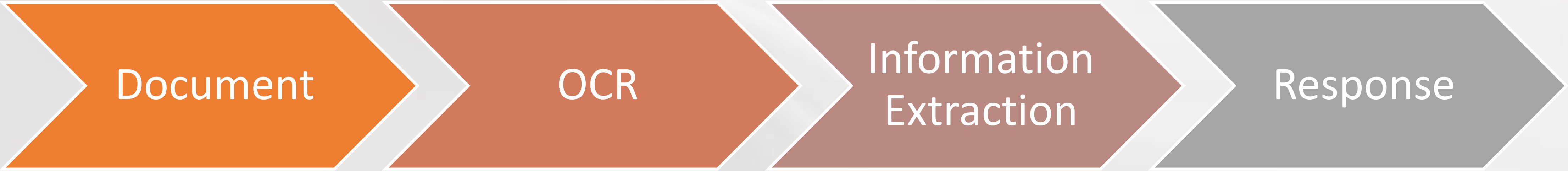

Task Dependent Outputs
Task Name:
Task Output
Word/Sentence BB Task:
X1 X2 Y1 Y2 </s>
BB-Word/Sentence Task:
Chicago </s>
Masked Document Text Generation:
Obama was a president of the United States of America. Obama was born in Honolulu. Hawaii. After Graduating </s>
QA Task:
Honolulu, Hawaii </s>
Obama was a [redacted] of the United States of America. Obama [redacted] born in Honolulu, Hawaii. After graduating from Columbia University in [redacted] he worked as a [redacted] organizer in Chicago.
Masked Text Input Image for Masked Document Generation Task
Obama was a president of the United States of America. Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago.
Unmasked Input Image for QA task and Bounding Box task
Image Encoder
Task Output
Text Decoder
Task Input
Task Name:
Task Input
Word/Sentence BB Task:
1983 [BB]
BB-Word/Sentence Task:
[BB] X1 X2 Y1 Y2 </s>
Masked Document Text Generation:
</s>
QA Task:
[QA] Where was Obama Born? </s>
Random sampling from these tasks at ith training step

Figure 2: Illustration of three tasks in the DUBLIN pretraining framework: Bounding Box, Rendered QA, and Masked Document Text Generation.

[2305.14218] DUBLIN -- Document Understanding By Language-Image Network (arxiv.org)

Traditional Document Processing Pipeline



Document

- Image extraction
- PDF Text extraction

OCR

- Image Pre-process
- Text Post-process

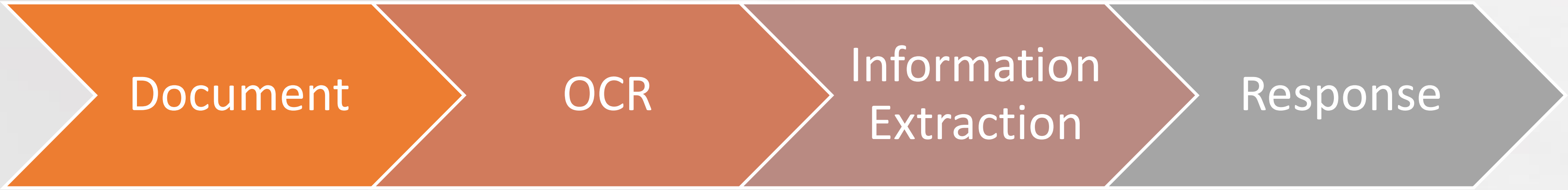
Information Extraction

- NER
- Results Post Process

Response

- Format
  - JSON

Modern Document Processing Pipeline



Document

OCR

Information Extraction

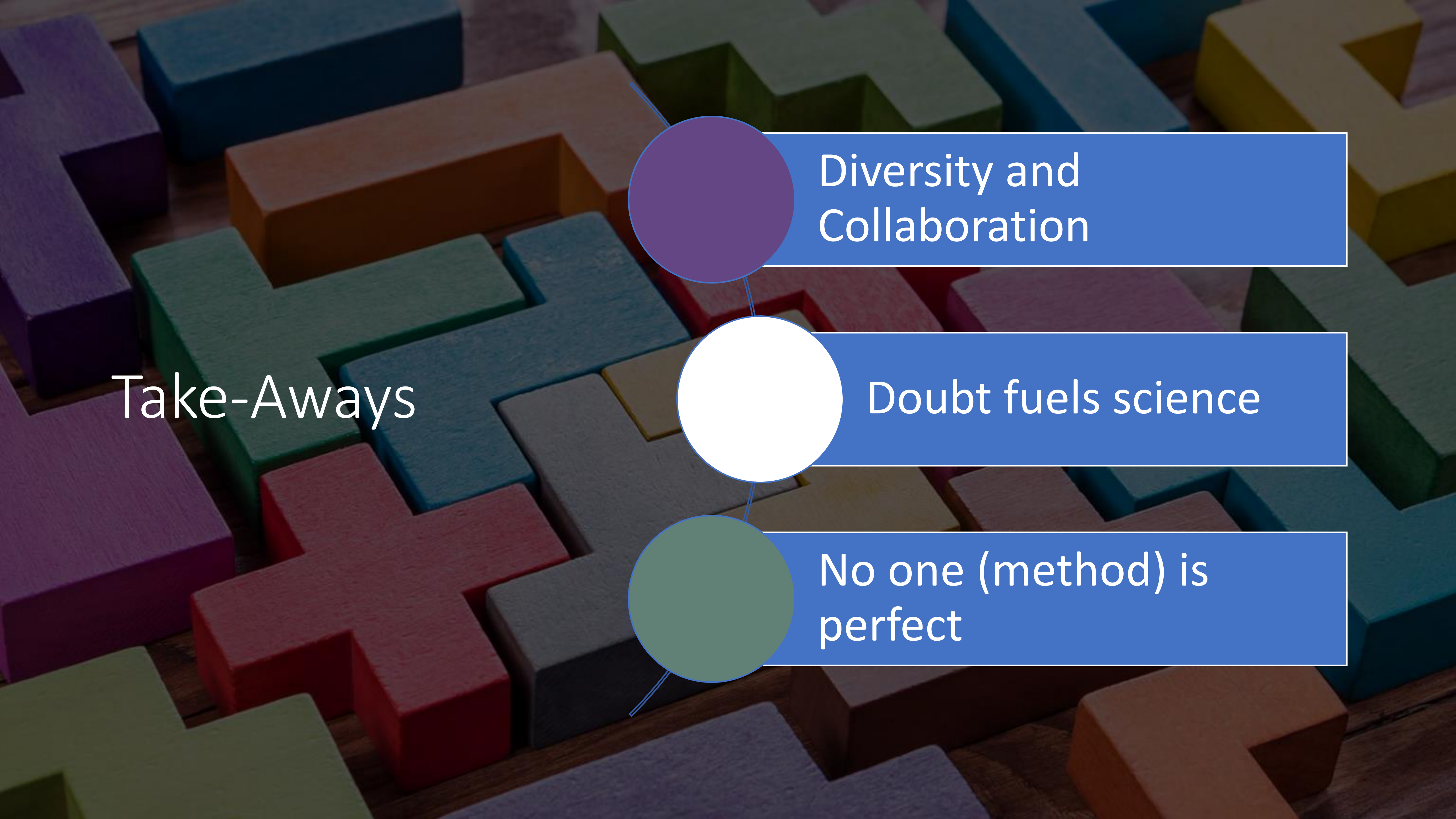
Response

- Image extraction
- PDF Text extraction

- Image Pre-process
- Text Post-process

- NER
- Results Post Process

- Format
  - JSON



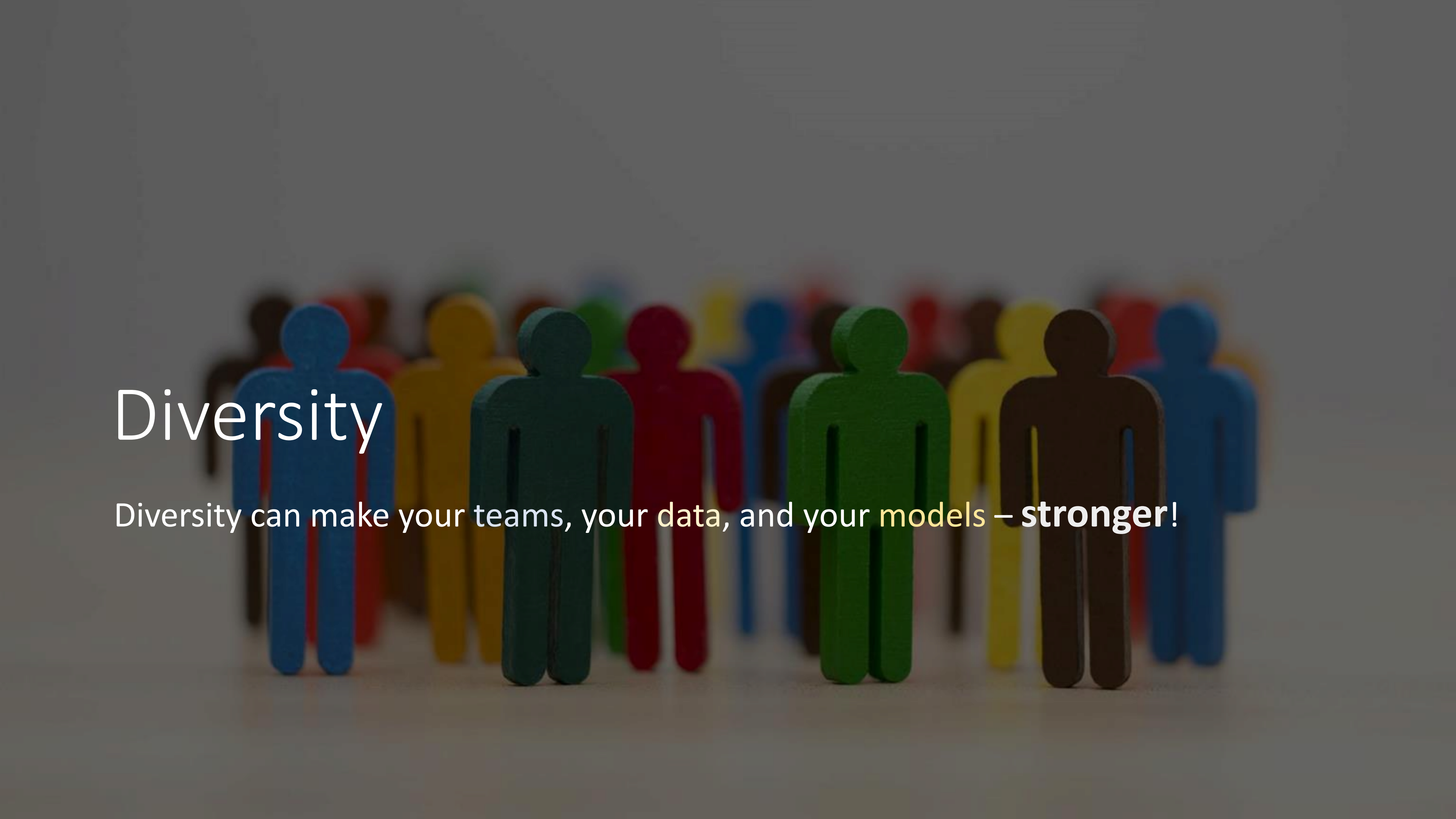
Take-Aways

- Diversity and Collaboration
- Doubt fuels science
- No one (method) is perfect

Thank you for your attention!

References

- The EU invoice volume and the scalability of our blockchain-based invoice reporting system · summitto blog
- sOCRates - a post-OCR text correction method | Anais do Simpósio Brasileiro de Banco de Dados (SBBD) (sbc.org.br)
- [2108.04539] BROS: A Pre-trained Language Model Focusing on Text and Layout for Better Key Information Extraction from Documents (arxiv.org)
- [2111.15664] OCR-free Document Understanding Transformer (arxiv.org)
- [2204.08387] LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking (arxiv.org)
- [2308.13418] Nougat: Neural Optical Understanding for Academic Documents (arxiv.org)
- [2203.16618] End-to-end Document Recognition and Understanding with Dessurt (arxiv.org)
- An Invoice Reading System Using a Graph Convolutional Network | SpringerLink



Diversity

Diversity can make your teams, your data, and your models – **stronger**!

International Conference on Document Analysis and Recognition (ICDAR)