Unifying Vision, Text, and Layout for Universal Document Processing

Zineng Tang[1,2], Ziyi Yang[2*], Guoxin Wang[3], Yuwei Fang[2], Yang Liu[2], Chenguang Zhu[2], Michael Zeng[2], Cha Zhang[3], Mohit Bansal[1*]

[1]University of North Carolina at Chapel Hill [2]Microsoft Azure Cognitive Services Research
[3]Microsoft Azure Visual Document Intelligence
*Corresp. authors: ziyiyang@microsoft.com, mbansal@cs.unc.edu

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

CVPR 2023 (Highlights)

Paper Tag: THU-AM-264 Paper ID: 8282

Agenda

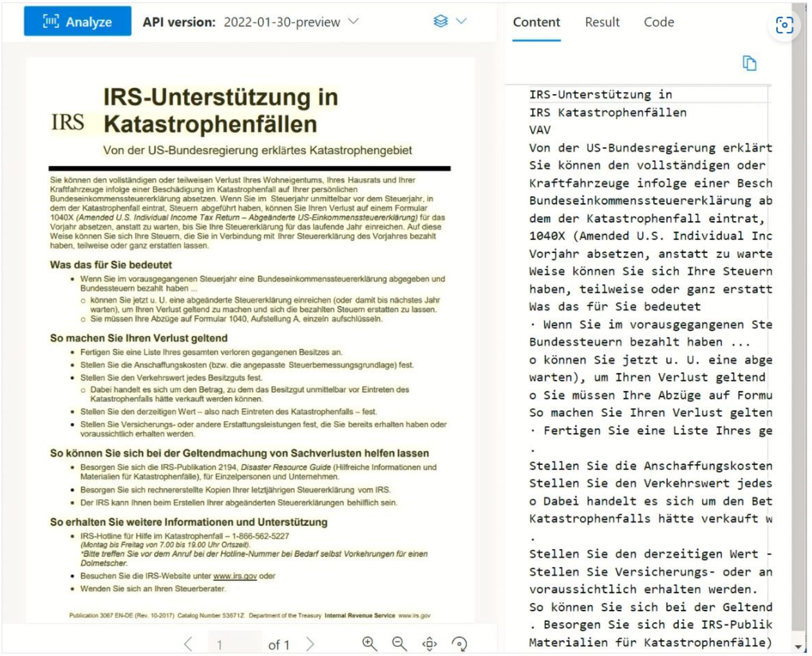
- Document AI Background, Challenges & Motivations
- UDOP Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

What is Document AI?

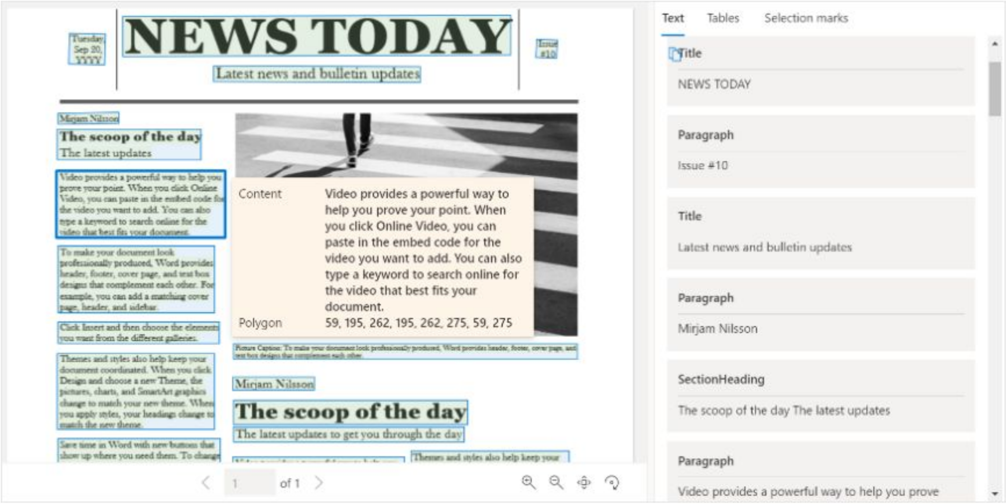
Parse, Analyze, and Understand Documents (receipt, paper, form, etc.)

Examples:

OCR



Layout Analysis



Document QA

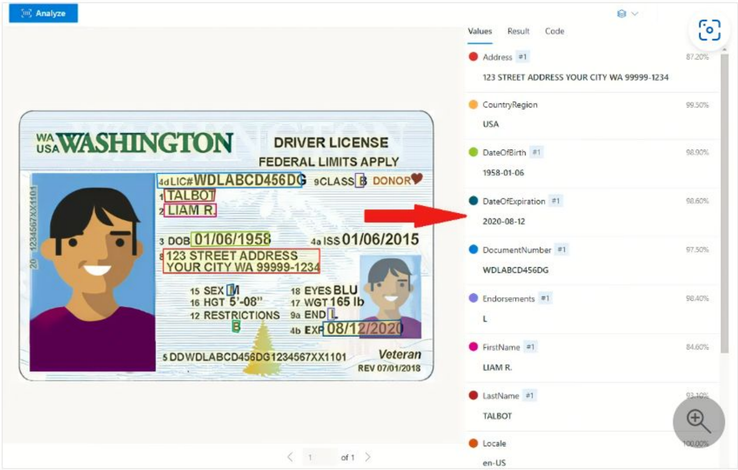
What is the interest rates of European Central Bank and US FED?

Image resource: [Document processing models - Form Recognizer - Azure Applied AI Services | Microsoft Learn](), [DocVQA]()

General Pipeline for Document AI

1. Convert **unstructured document** into **structured data**

- Optical Character Recognition
- PDF converter
- HTML reader
- …



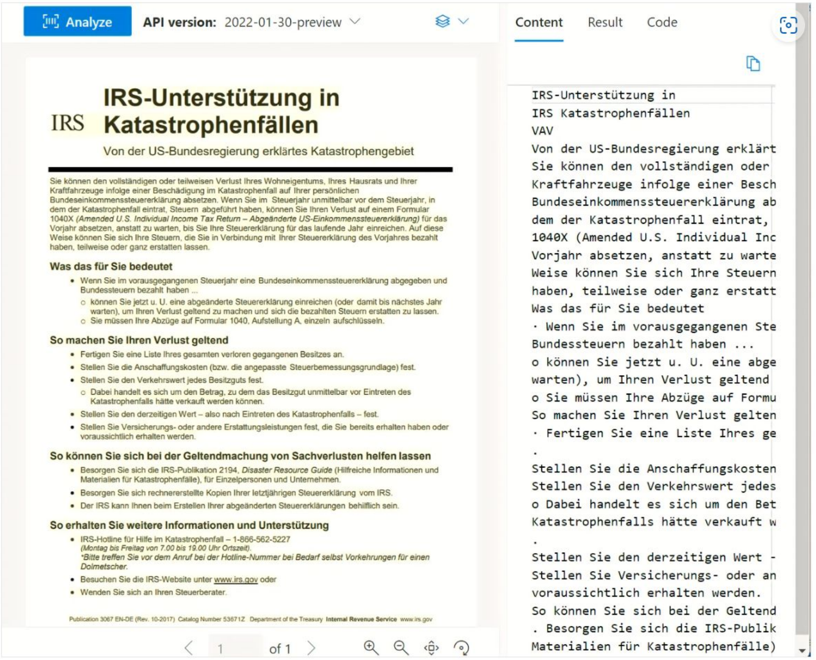
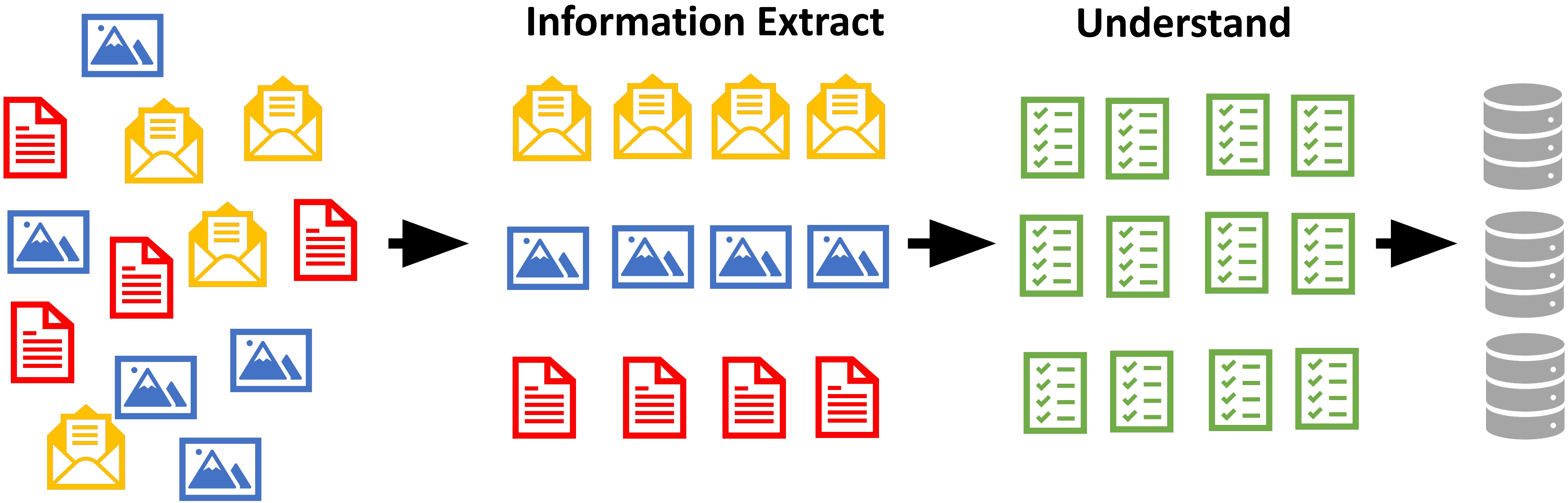


Image resource: [Document processing models - Form Recognizer - Azure Applied AI Services | Microsoft Learn]

General Pipeline for Document AI

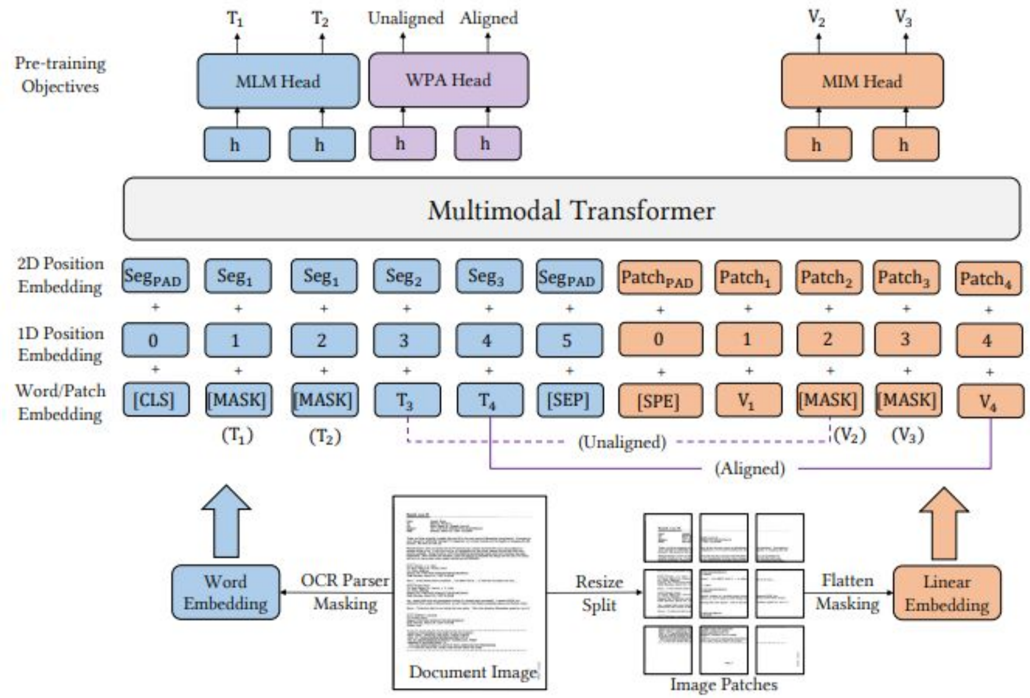
2. Extract, Classify & Understand Information from Document



Adapted from [What is Document AI? - YouTube]

Previous Works

LayoutLM v3



Donut

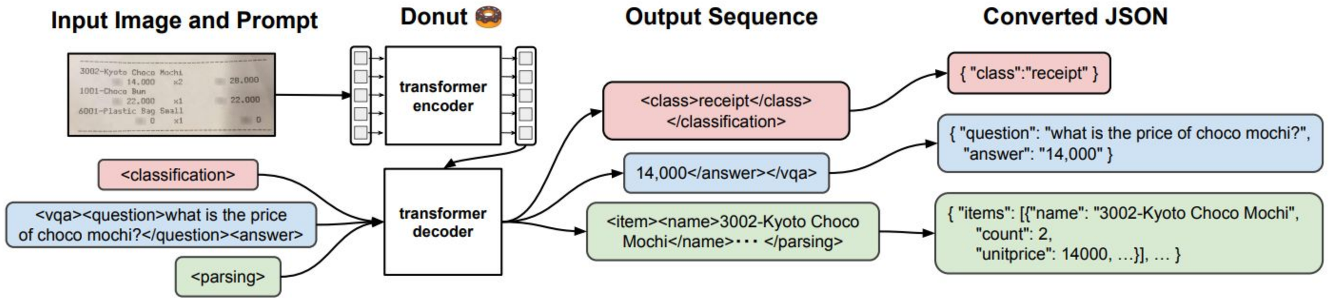


Image source: [LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking (arxiv.org)]
[2111.15664] OCR-free Document Understanding Transformer (arxiv.org)

Challenges in Document AI & Motivations

1. The correlation between image and text is very strong in document data

Classical Vision-Language
text are usually high-level description
of the vision data

What color are her eyes?
What is the mustache made of?

Challenges in Document AI & Motivations

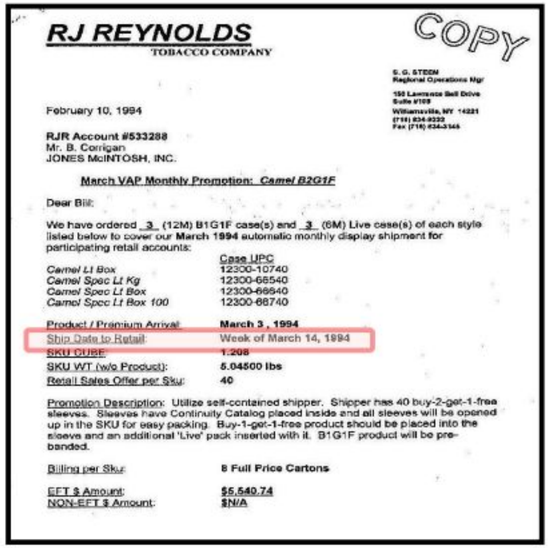
1. The correlation between image and text is very strong in document data

Classical Vision-Language
text are usually high-level description
of the vision data

What color are her eyes?
What is the mustache made of?

V.S.

Document AI
1-to-1 correspondence between text
tokens and image regions



RJ REYNOLDS
TOBACCO COMPANY

COPY

February 10, 1994

RJR Account #533288
Mr. B. Corrigan
JONES McINTOSH, INC.

March VAP Monthly Promotion: *Camel B2G1F*

Dear Bill:

We have ordered 3 (12M) B1G1F case(s) and 3 (6M) Live case(s) of each style listed below to cover our **March 1994** automatic monthly display shipment for participating retail accounts:

	Case UPC
Camel Lt Box	12300-10740
Camel Spec Lt Kg	12300-66540
Camel Spec Lt Box	12300-66640
Camel Spec Lt Box 100	12300-66740

Product / Premium Arrival:	**March 3 , 1994**
Ship Date to Retail:	Week of March 14, 1994
SKU CUBE:	**1.208**
SKU WT (w/o Product):	**5.04500 lbs**
Retail Sales Offer per Sku:	**40**

Promotion Description: Utilize self-contained shipper. Shipper has 40 buy-2-get-1-free sleeves. Sleeves have Continuity Catalog placed inside and all sleeves will be opened up in the SKU for easy packing. Buy-1-get-1-free product should be placed into the sleeve and an additional 'Live' pack inserted with it. B1G1F product will be pre-banded.

Billing per Sku:	**8 Full Price Cartons**
EFT $ Amount:	**$5,540.74**
NON-EFT $ Amount:	**$N/A**

Ship	*(0.1, 0.65, 0.11, 0.68)*
Date	*(0.12, 0.65, 0.14, 0.68)*
to	*(0.16, 0.65, 0.18, 0.68)*

Challenges in Document AI & Motivations

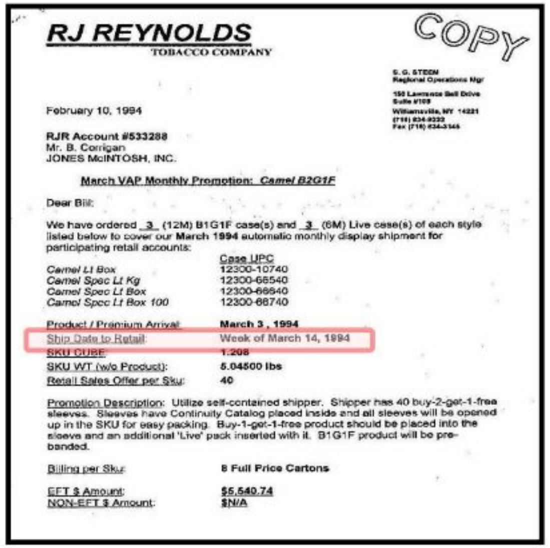
1. The correlation between image and text is very strong in document data

Classical Vision-Language
text are usually high-level description of the vision data

What color are her eyes?
What is the mustache made of?

V.S.

Document AI
1-to-1 correspondence between text tokens and image regions



RJ REYNOLDS
TOBACCO COMPANY

COPY

February 10, 1994

RJR Account #533288
Mr. B. Corrigan
JONES McINTOSH, INC.

March VAP Monthly Promotion: *Camel B2G1F*

Dear Bill:

We have ordered 3 (12M) B1G1F case(s) and 3 (6M) Live case(s) of each style listed below to cover our **March 1994** automatic monthly display shipment for participating retail accounts:

	Case UPC
Camel Lt Box	12300-10740
Camel Spec Lt Kg	12300-66540
Camel Spec Lt Box	12300-66640
Camel Spec Lt Box 100	12300-66740

Product / Premium Arrival:	**March 3 , 1994**
Ship Date to Retail:	Week of March 14, 1994
SKU CUBE:	**1.208**
SKU WT (w/o Product):	**5.04500 lbs**
Retail Sales Offer per Sku:	**40**

Promotion Description: Utilize self-contained shipper. Shipper has 40 buy-2-get-1-free sleeves. Sleeves have Continuity Catalog placed inside and all sleeves will be opened up in the SKU for easy packing. Buy-1-get-1-free product should be placed into the sleeve and an additional 'Live' pack inserted with it. B1G1F product will be pre-banded.

Billing per Sku:	**8 Full Price Cartons**
EFT $ Amount:	$5,540.74
NON-EFT $ Amount:	$N/A

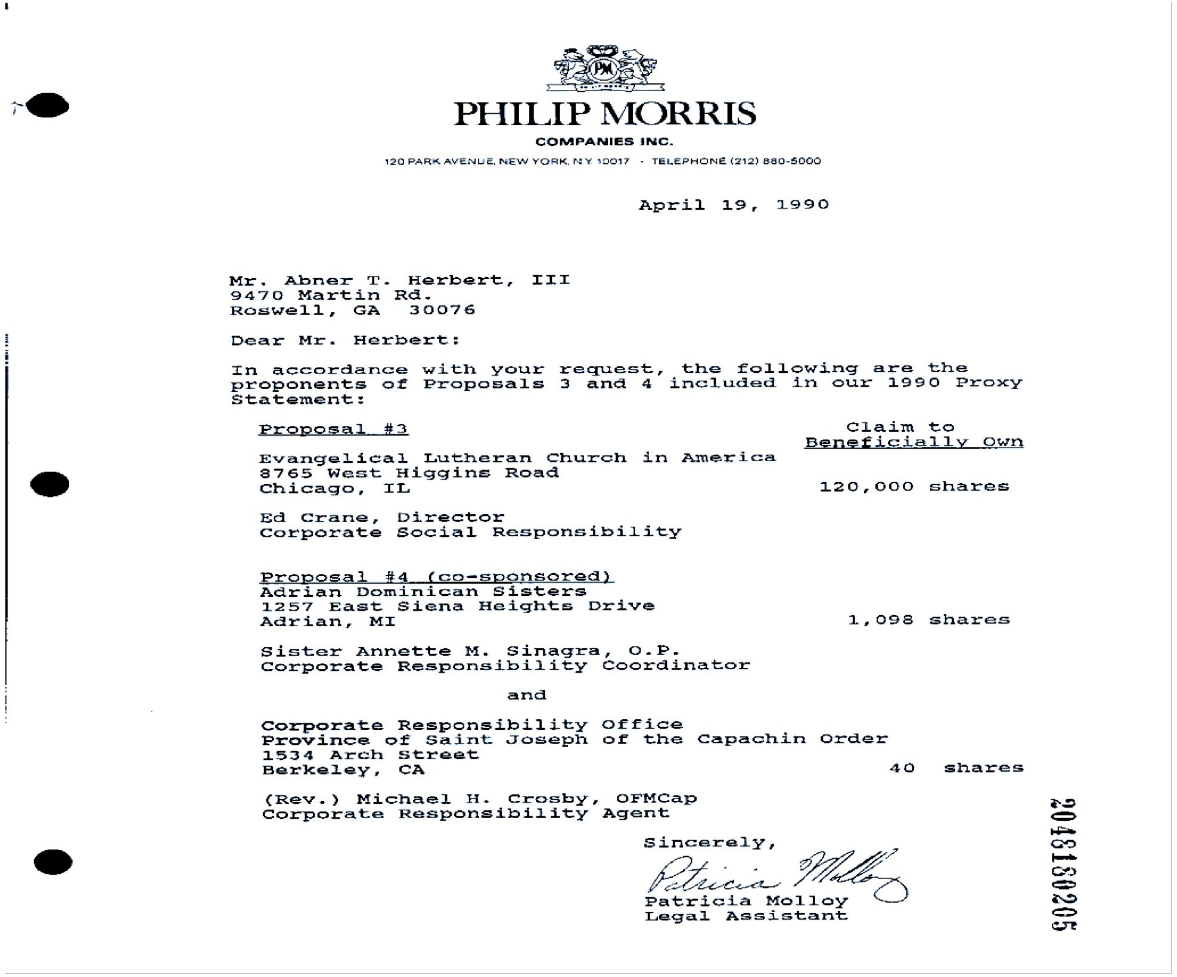
Ship	*(0.1, 0.65, 0.11, 0.68)*
Date	*(0.12, 0.65, 0.14, 0.68)*
to	*(0.16, 0.65, 0.18, 0.68)*

Previous Document AI works

- Mostly use 1D or 2D positional embeddings
- Indirect & Insufficient

Challenges in Document AI & Motivations

2. Document AI tasks are diverse



PHILIP MORRIS
COMPANIES INC.
120 PARK AVENUE, NEW YORK, N.Y 10017 · TELEPHONE (212) 880-5000

April 19, 1990

Mr. Abner T. Herbert, III
9470 Martin Rd.
Roswell, GA 30076

Dear Mr. Herbert:

In accordance with your request, the following are the proponents of Proposals 3 and 4 included in our 1990 Proxy Statement:

Proposal #3	Claim to Beneficially Own
Evangelical Lutheran Church in America 8765 West Higgins Road Chicago, IL	120,000 shares
Ed Crane, Director Corporate Social Responsibility	
Proposal #4 (co-sponsored) Adrian Dominican Sisters 1257 East Siena Heights Drive Adrian, MI	1,098 shares
Sister Annette M. Sinagra, O.P. Corporate Responsibility Coordinator	
and	
Corporate Responsibility Office Province of Saint Joseph of the Capachin Order 1534 Arch Street Berkeley, CA	40 shares
(Rev.) Michael H. Crosby, OFMCap Corporate Responsibility Agent	

Sincerely,

Patricia Molloy
Patricia Molloy
Legal Assistant

2048180205

Document Classification:
What is the type of the document?

Document QA:
What is the address of Philip Morris Companies Inc?

Layout Detection:
Where is the signature?

Information Extraction:
Document serial number: 2048180205

Customized Document Image Generation

…

Previous approaches:
Need a **different** model head for **each** task.

Challenges in Document AI & Motivations

3. Self-supervised pretraining tasks in previous works were not specifically designed for Document AI.

- From single modality pretraining:

- Masked Language Modeling (text)
- Masked Image Modeling (vision)

- From classical VL training:

- Text-Vision contrastive learning

Challenges in Document AI & Motivations

4. Previous works only used unlabeled document data*. Diverse and abundant supervised data are ignored.

PHILIP MORRIS
COMPANIES INC.
120 PARK AVENUE, NEW YORK, N.Y. 10017 · TELEPHONE (212) 880-5000

April 19, 1990

Mr. Abner T. Herbert, III
9470 Martin Rd.
Roswell, GA 30076

Dear Mr. Herbert:

In accordance with your request, the following are the proponents of Proposals 3 and 4 included in our 1990 Proxy Statement:

Proposal #3	Claim to Beneficially Own
Evangelical Lutheran Church in America 8765 West Higgins Road Chicago, IL	120,000 shares
Ed Crane, Director Corporate Social Responsibility	
Proposal #4 (co-sponsored) Adrian Dominican Sisters 1257 East Siena Heights Drive Adrian, MI	1,098 shares
Sister Annette M. Sinagra, O.P. Corporate Responsibility Coordinator	
and	
Corporate Responsibility Office Province of Saint Joseph of the Capachin Order 1534 Arch Street Berkeley, CA	40 shares
(Rev.) Michael H. Crosby, OFMCap Corporate Responsibility Agent	

Sincerely,
Patricia Molloy
Patricia Molloy
Legal Assistant

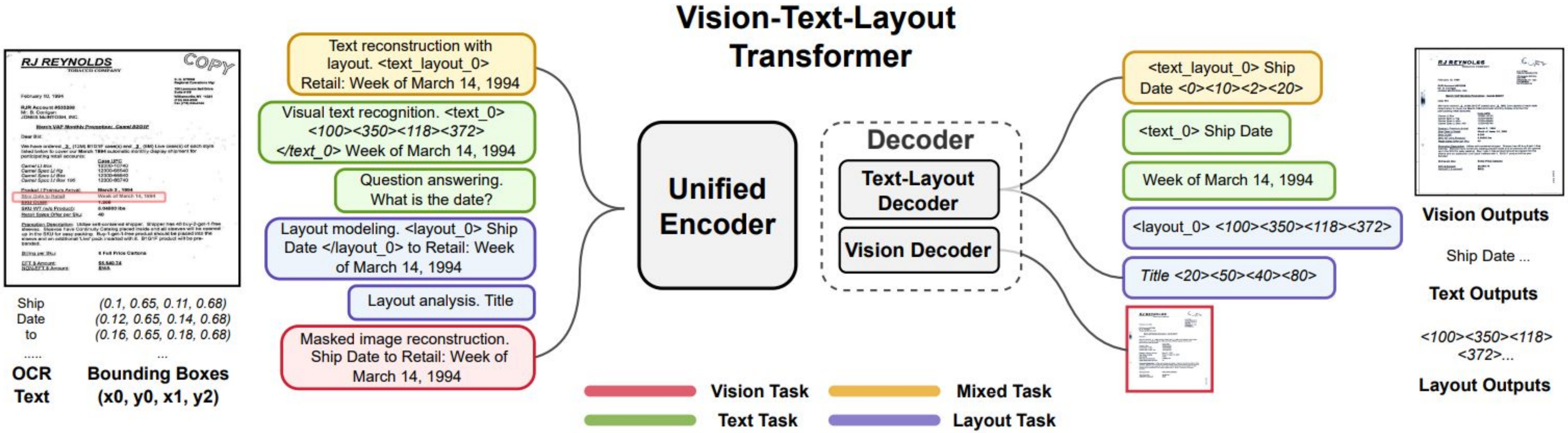
2048180205

Supervised Data | **Dataset**

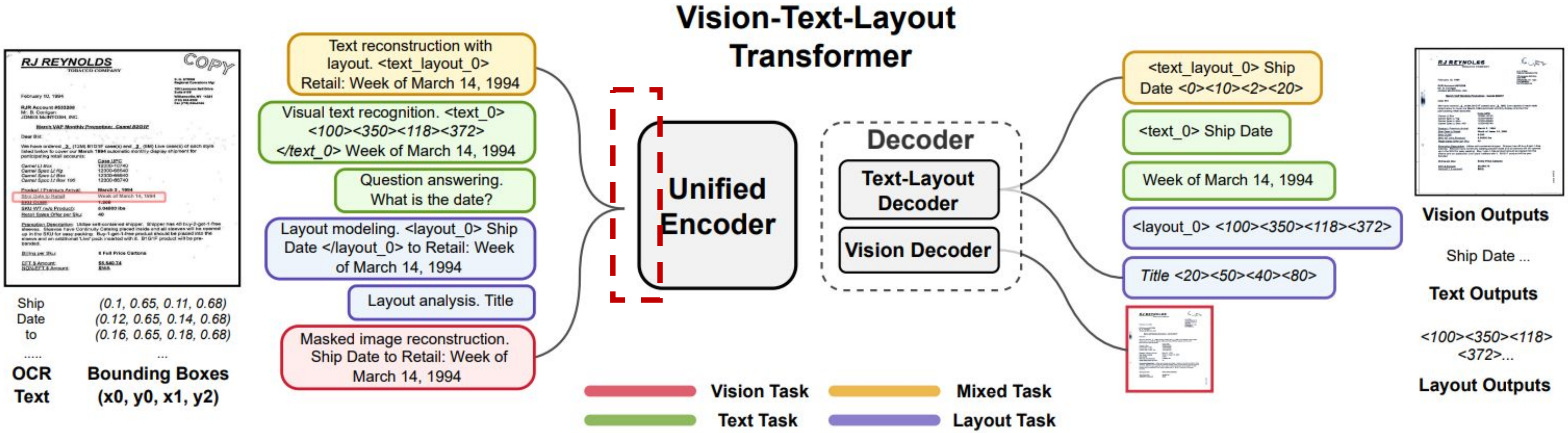
Supervised Data	Dataset
Document Classification: What is the type of the document?	RVL-CDIP
Document QA: What is the address of Philip Morris Companies Inc?	WebSRC, VisualMRC, DocVQA, InfographicsVQA, WTQ
Layout Detection: Where is the signature?	PublayNet
Information Extraction: Document serial number: 2048180205	DocBank, KLC, PWC, DeepForm
Document NLI: Predict the "entailment" or not between a sentence pair given a document	TabFact

*Some works, e.g. LayoutLM, use one anxuliarry task in pretraining

UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing

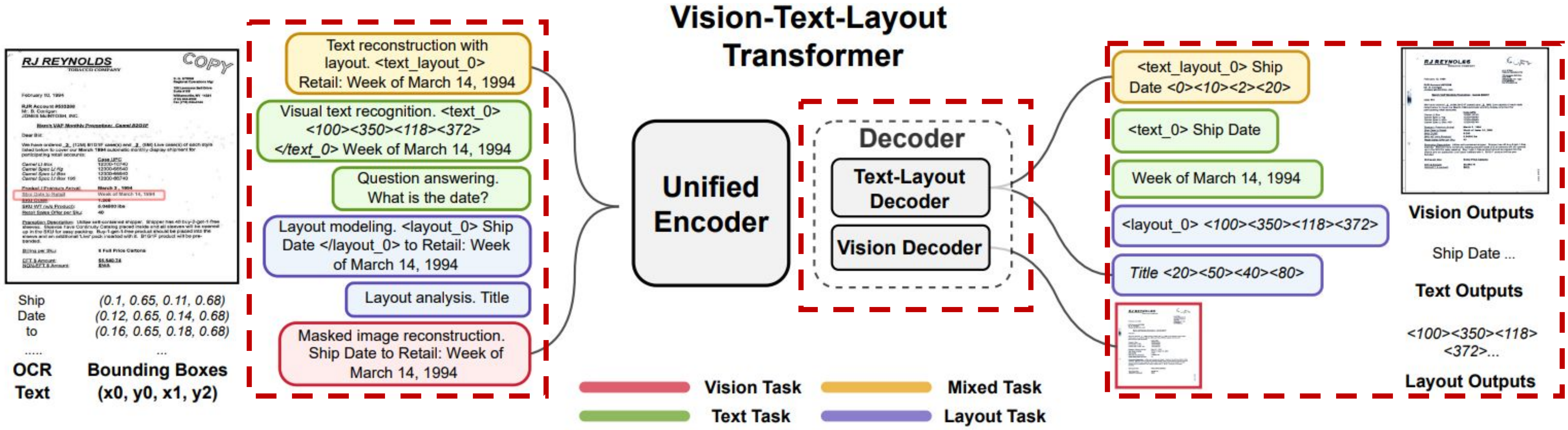


UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing



1. The layout-induced vision-text embedding: leveraging the strong correlation between text and vision modality.

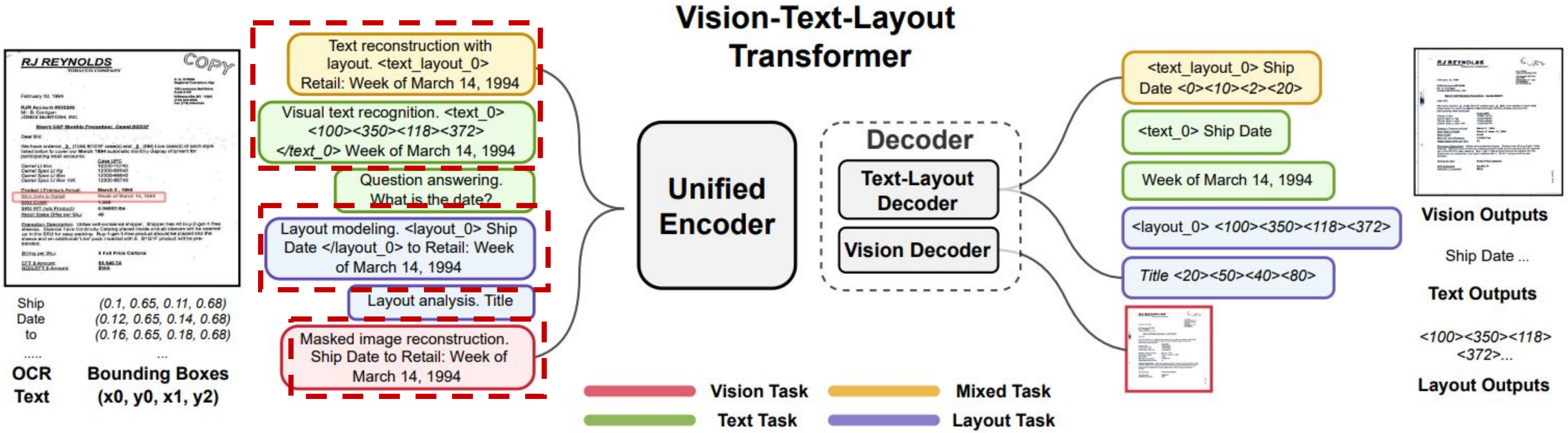
UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing



2. UDOP (i-Code Doc) is a generative framework

- Can model all existing document **understanding** and document **generation** tasks with **one unified model**

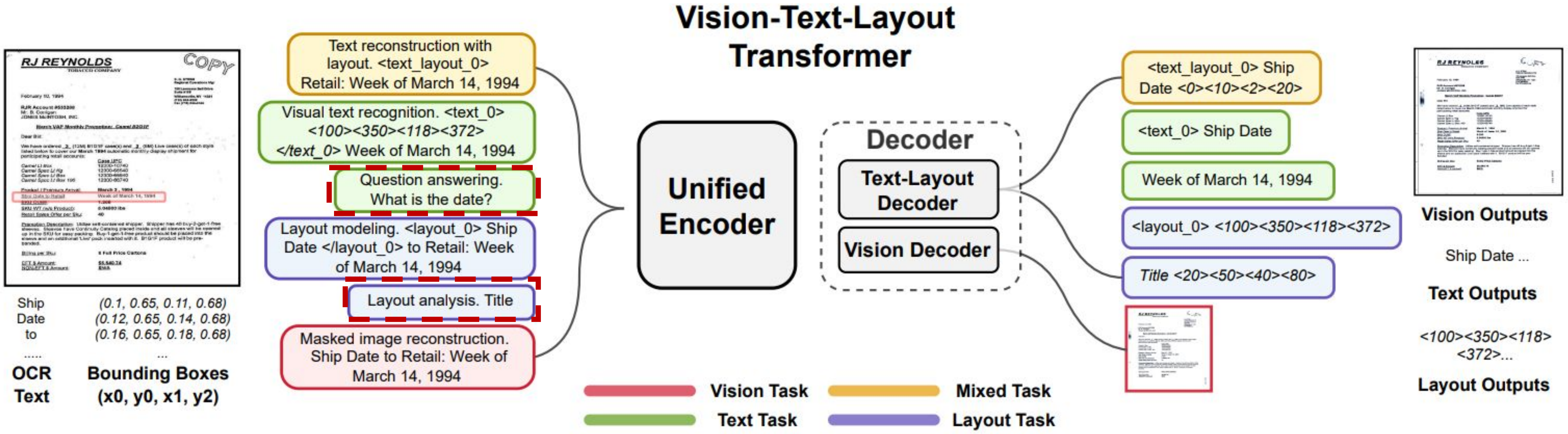
UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing



3. Novel pretraining framework

- **Self-supervised** pretraining objectives **specifically designed for Document AI**

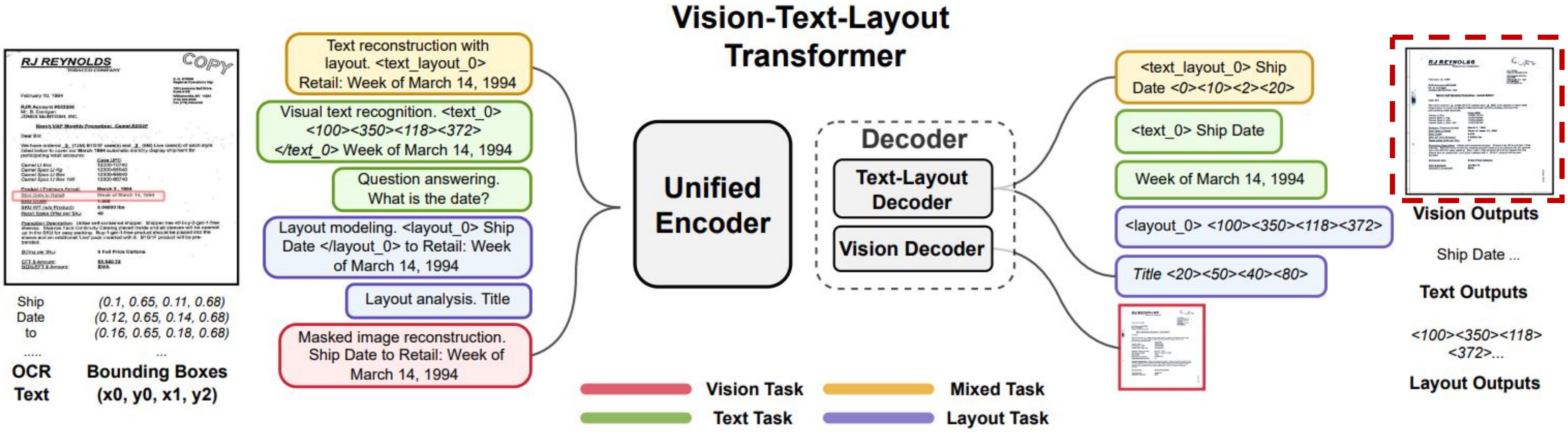
UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing



3. Novel pretraining framework

- **Self-supervised** pretraining objectives **specifically designed for Document AI**
- Included **previously-ignored supervised data** for pretraining

UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing

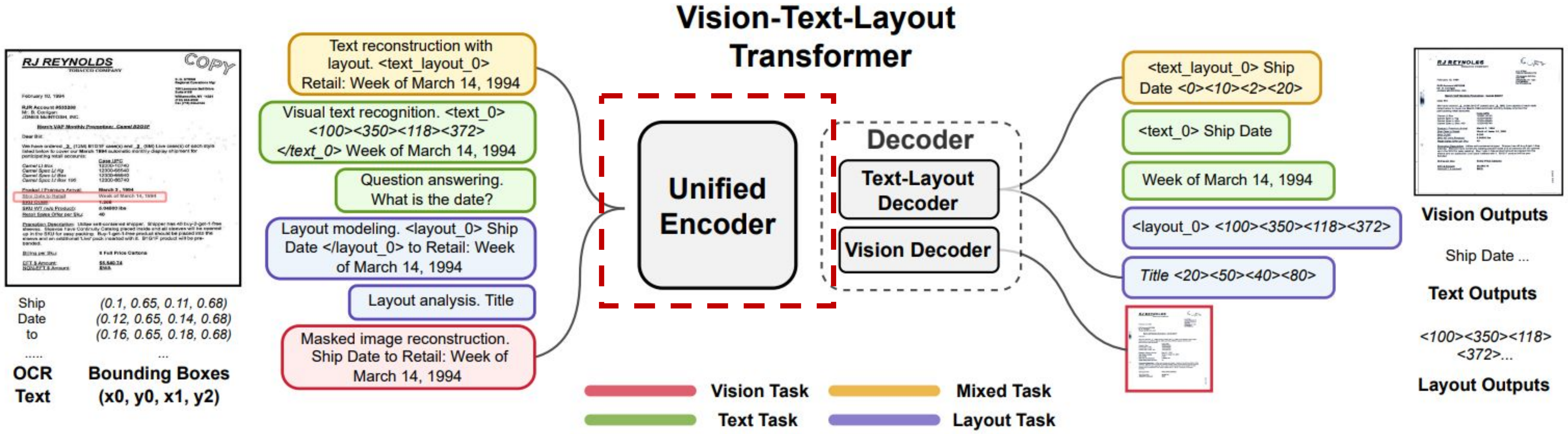


4. High-quality controllable **document generation**

Agenda

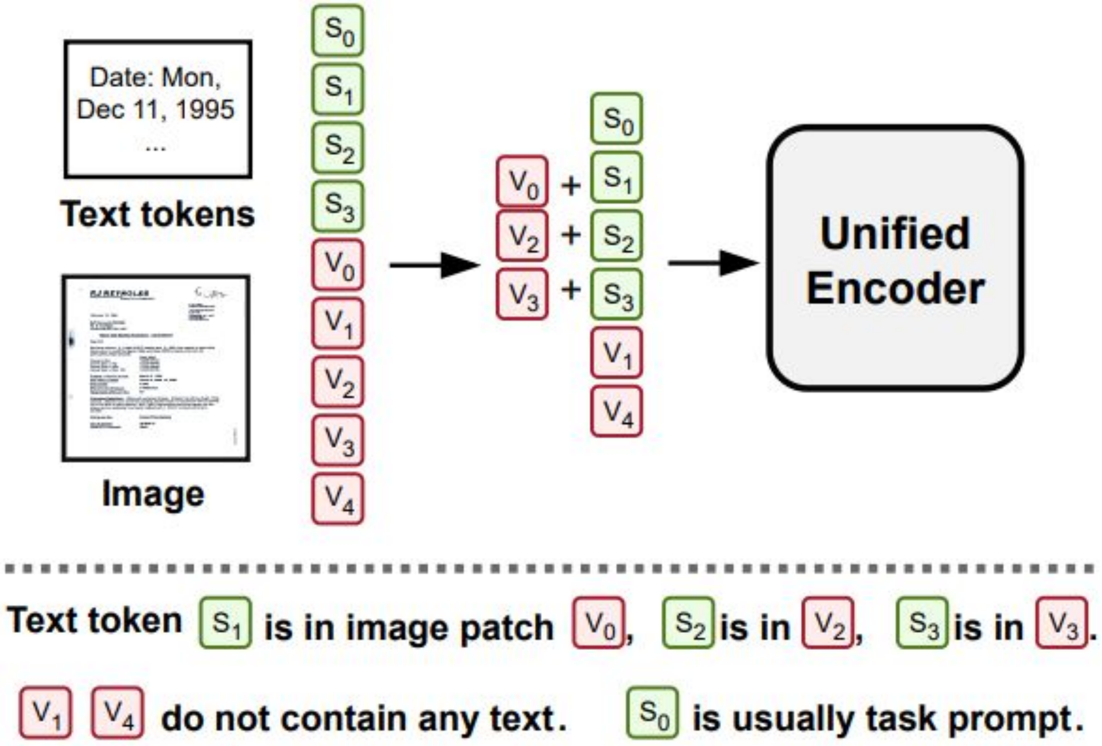
- Document AI Background, Challenges & Motivations
- Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

UDOP (i-Code Doc): Unifying Vision, Text, and Layout for Universal Document Processing



A Unified Vision, Text, and Layout Encoder

Layout-Induced Vision-Text Embedding



Definition of Notations

Concretely, given the document image $\boldsymbol{v} \in \mathbb{R}^{H \times W \times C}$, M word tokens $\{s_i\}_{i=1}^{M}$ inside the image and the extracted layout structure $\{(x_i^1, y_i^1, x_i^2, y_i^2)\}_{i=1}^{M}$, we first partition $\boldsymbol{v}$ into $\frac{H}{P} \times \frac{W}{P}$ image patches, where each patch is of size $P \times P \times C$. We then encode each patch with a D-dim vector and group all patch embeddings into a sequence of vectors $\{\boldsymbol{v}_i \in \mathbb{R}^D\}_{i=1}^{N}$ where $N = \frac{H}{P} \times \frac{W}{P}$. Text tokens are also converted to numerical D-dim embeddings $\{\boldsymbol{s}_i\}_{i=1}^{M}$ by vocabulary look-up.

Or formally, first define the layout indicator function

$$\phi(\boldsymbol{s}_i, \boldsymbol{v}_j) = \begin{cases} 1, & \text{if the center of } \boldsymbol{s}_i\text{'s bounding box} \\ & \text{is within the image patch } \boldsymbol{v}_j. \\ 0, & \text{otherwise.} \end{cases}$$

Then for each text token embedding $\boldsymbol{s}_i$, the joint representation is the sum of its image patch feature[2] and the text feature:

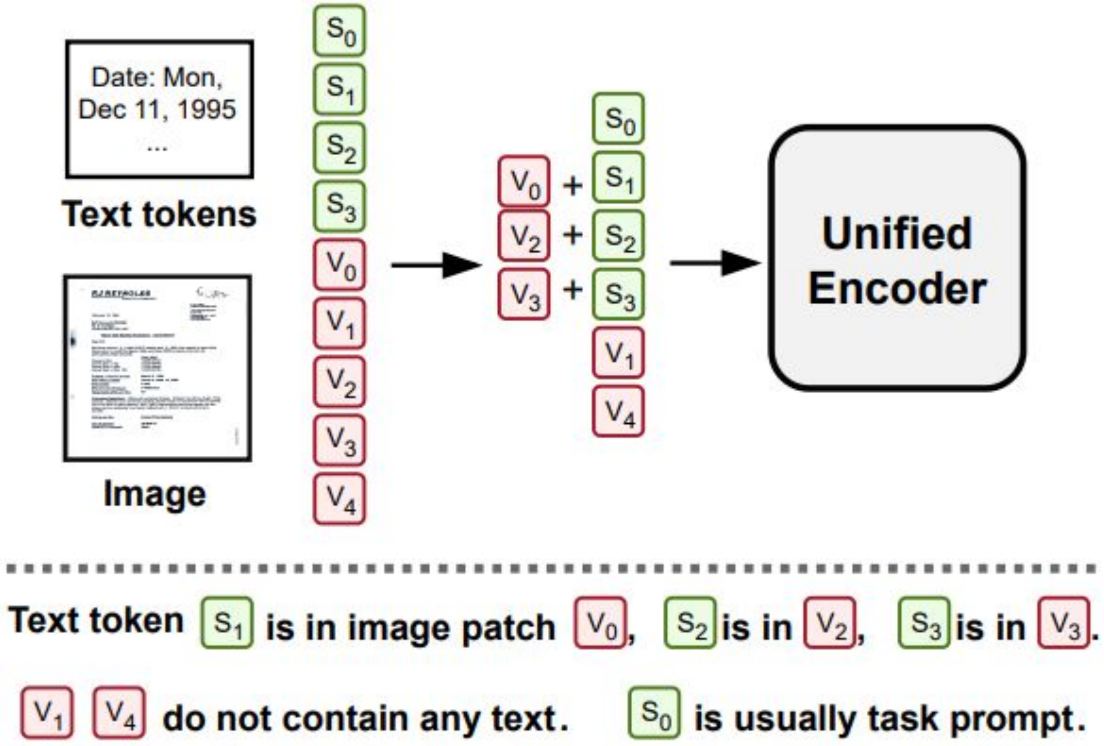
$$\boldsymbol{s}_i' = \boldsymbol{s}_i + \boldsymbol{v}_j, \text{ where } \phi(\boldsymbol{s}_i, \boldsymbol{v}_j) = 1.$$

For image patches $\boldsymbol{v}_j$ without any text tokens, i.e. $\forall i,\ \phi(\boldsymbol{s}_i, \boldsymbol{v}_j) = 0$, the joint representation, $\boldsymbol{v}_j'$ is itself:

$$\boldsymbol{v}_j' = \boldsymbol{v}_j.$$

A Unified Vision, Text, and Layout Encoder

Layout-Induced Vision-Text Embedding



Layout Tokens

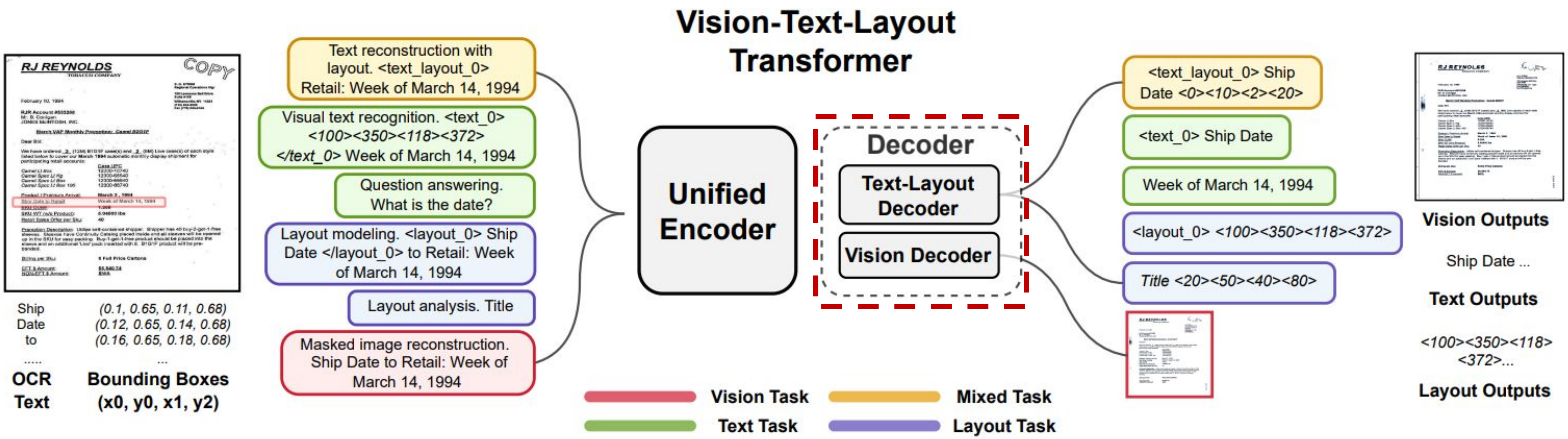
Convert bounding box coordinates to discretized tokens

$(0.1, 0.2, 0.5, 0.6)$ □ *<50><100><250><300>*

(Assuming layout vocab size 500)

- Convenient for location detection tasks: layout token generation
- Integrating layout modeling in pretraining

Vision, Text, and Layout Decoder



Text-Layout Decoder: Generate textual sequence/layout tokens
Vision Decoder: Generate document images

Agenda

- Document AI Background, Challenges & Motivations
- Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

Pretraining

Self-supervised Pretraining

Use “Ship Date to Retail: Week of March 14, 1994” as example

(1) Joint Text-Layout Reconstruction

Input Sequence: “*Joint Text-Layout Reconstruction*. <text_layout_0> to Retail: Week <text_layout_1> March 14, 1994”
Target Sequence: “<text_layout_0> Ship Date *<100><350><118><372>* <text_layout_1> of *<100><370><118><382>*”

(2) Layout Modeling

Input Sequence: “*Layout Modeling*. <layout_0> Ship Date </layout_0> to Retail: Week <layout_1> of </layout_1> March 14, 1994”
Target Sequence: “<layout_0> *<100><350><118><372>* <layout_1> *<100><370><118><382>*”

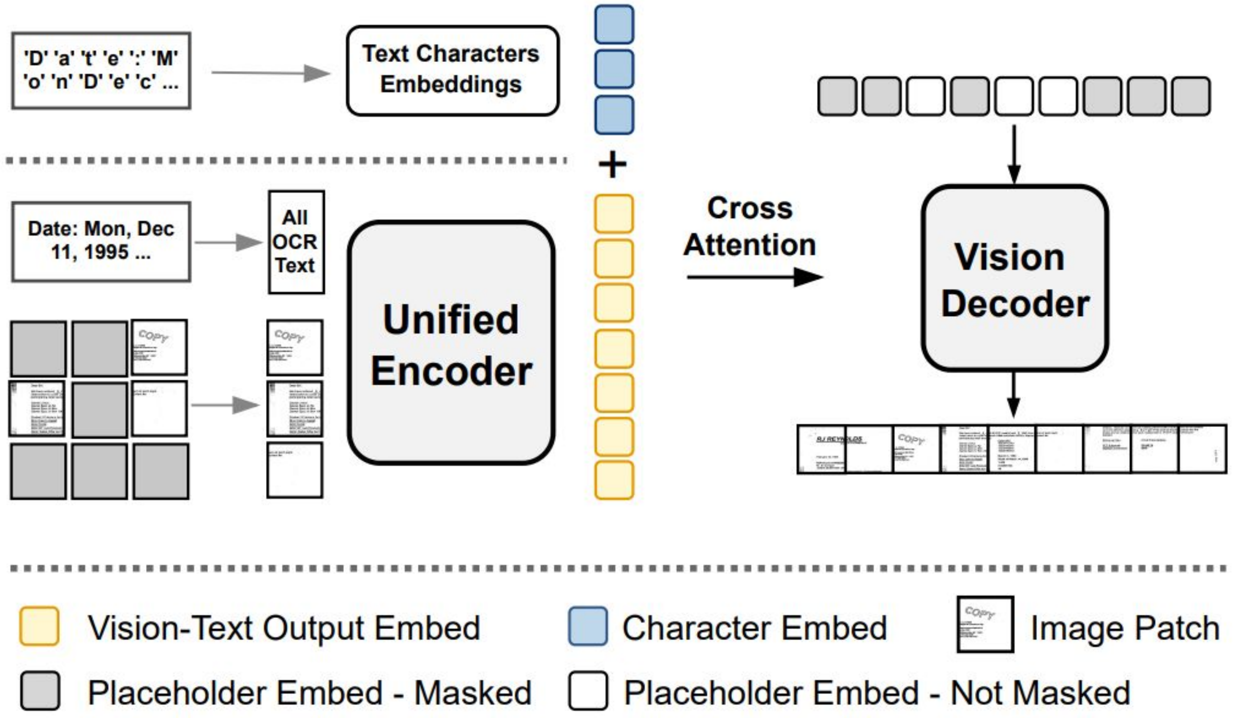
(3) Visual Text Recognition

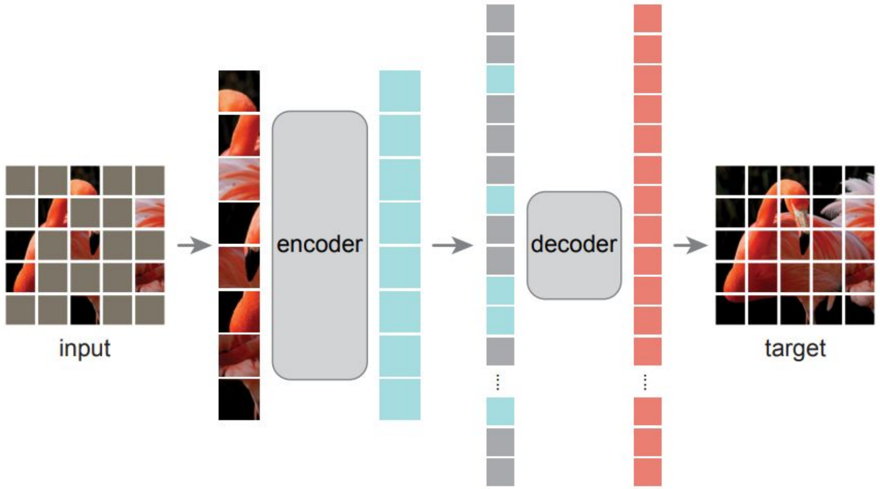
Input Sequence: “*Visual Text Recognition*. <text_0> *<100><350><118><372>* </text_0> to Retail: Week <text_1> *<100><370><118><382>* </text_1> March 14, 1994”
Target Sequence: “<text_0> Ship Date <text_1> of”

Pretraining

Self-supervised Pretraining

(4) Masked Image Reconstruction with Text and Layout





The original Masked AutoEncoder (MAE) design (He et al.)

Pretraining

Supervised Pretraining: Unifying All Tasks into the Generative Scheme

Supervised Tasks	**Task Prompts**	**Task Targets**
Classification	*Document Classification*. Ship Date to Retail: Week of March 14, 1994	Memo.
Layout Analysis	*Layout Analysis*. Paragraph.	Paragraph <*82*><*35*><*150*><*439*>
Information Extraction	*Information Extraction*. Ship Date to Retail	Week of March 14, 1994
Question Answering	*Question Answering*. What is the ship year?	1994
Document NLI	*Document Natural Language Inference*. Ship Date to Retail: Week of March 14, 1994	Entailment.

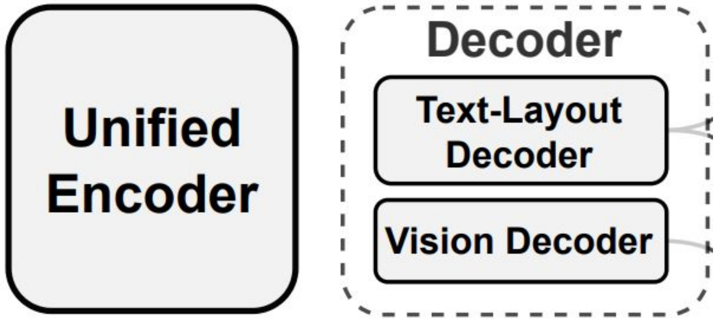
Pretraining

Supervised Pretraining: datasets

Supervised Data	Dataset
Document Classification: What is the type of the document?	RVL-CDIP
Document QA: What is the address of Philip Morris Companies Inc?	WebSRC, VisualMRC, DocVQA, InfographicsVQA, WTQ
Layout Detection: Where is the signature?	PublayNet
Information Extraction: Document serial number: 2048180205	DocBank, KLC, PWC, DeepForm
Document NLI: Predict the “entailment” or not between a sentence pair given a document	TabFact

Pretraining

Model Configuration:



- Unified Encoder + Text Layout Decoder: T5 large
- Vision Decoder: MAE-large
- 794M parameters

Curriculum Learning:

Image resolution 224 □ 512 □ 1024

Train with each resolution with 1 epoch

Agenda

- Document AI Background, Challenges & Motivations
- i-Code Doc Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

Evaluation

	Model	Modality	Info Ext.		Classification
			FUNSD	CORD	RVL-CDIP
NAVER	Donut [21]	V	-	91.6	95.3
Google	$\text{BERT}_{\text{large}}$ [9]	T	65.63	90.25	89.92
	$\text{BROS}_{\text{large}}$ [15]	T+L	84.52	97.40	-
Alibaba Group	$\text{StructuralLM}_{\text{large}}$	T+L	85.14	-	**96.08**
CS CAMSCANNER	LiLT [48]	T+L	88.41	96.07	95.68
Google	FormNet [24]	T+L	84.69	97.28	-
MSRA	$\text{LayoutLM}_{\text{large}}$ [53]	T+L	77.89	-	91.90
Adobe	SelfDoc [29]	V+T+L	83.36	-	92.81
Adobe	UniDoc [11]	V+T+L	87.93	96.86	95.05
aws	$\text{DocFormer}_{\text{large}}$ [1]	V+T+L	84.55	96.99	95.50
	$\text{TILT}_{\text{large}}$ [36]	V+T+L	-	96.33	95.52
MSRA	$\text{LayoutLMv2}_{\text{large}}$ [55]	V+T+L	84.20	96.01	95.64
MSRA	$\text{LayoutLMv3}_{\text{large}}$ [16]	V+T+L	**92.08**	97.46	95.93
	UDOP	V+T+L	91.62	**97.58**	96.00

FUNSD (Form Understanding in Noisy Scanned Documents [18]) has 149 and 50 samples for train and test. We evaluate on the entity recognition task: predicting the entity, "question", "answer", "header", or "other", for the text token. The task format is, suppose we have the title, "The Title", and its entity "[I-Header]", then the encoder input is "The Title" and the generation target is "The Title [I-Header]". The metric is F1 scores.

CORD (Consolidated Receipt Dataset for Post-OCR Parsing) [33] is a key information extraction dataset with 30 labels under 4 categories such as "total" or "subtotal". It has 1,000 receipt samples. The train, validation, and test splits contain 800, 100, and 100 samples respectively. The metric is F1 and the task format is the same as FUNSD.

RVL-CDIP is the document classification dataset that we have discussed previously. It has 320k/40k/40k images for training/validation/test. The metric is classification accuracy.

Evaluation

	Model	Modality	Question Answering		Information Extraction			Table QA/NLI		Avg.
			DocVQA	InfoVQA	KLC	PWC	DeepForm	WTQ	TabFact	
NAVER	Donut [21]	V	72.1	-	-	-	-	-	-	-
Google	BERT$_{large}$ [9]	T	67.5	-	-	-	-	-	-	-
Google	T5$_{large}$ [39]	T	70.4	36.7	74.3	25.3	74.4	33.3	58.9	50.7
	T5$_{large}$+U [36]	T	76.3	37.1	76.0	27.6	82.9	38.1	76.0	56.5
	T5$_{large}$+2D [36]	T+L	69.8	39.2	72.6	25.7	74.0	30.8	58.0	50.4
	T5$_{large}$+2D+U [36]	T+L	81.0	46.1	75.9	26.8	83.3	43.3	78.6	59.8
	LAMBERT [10]	T+L	-	-	81.3	-	-	-	-	-
Alibaba Group	StructuralLM$_{large}$ [26]	T+L	83.9	-	-	-	-	-	-	-
MSRA	LayoutLMv2$_{large}$ [55]	V+T+L	78.8	-	-	-	-	-	-	-
MSRA	LayoutLMv3$_{large}$ [16]	V+T+L	83.4	45.1	77.1	26.9	84.0	45.7	78.1	62.9
	UDOP	V+T+L	**84.7**	**47.4**	**82.8**	**28.0**	**85.5**	**47.2**	**78.9**	**64.8**

Agenda

- Document AI Background, Challenges & Motivations
- Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

Document image editing & customization with i-Code Doc

Original Document

PHILIP MORRIS
COMPANIES INC.
120 PARK AVENUE, NEW YORK, N.Y 10017 - TELEPHONE (212) 880-5000

April 19, 1990

Mr. Abner T. Herbert, III
9470 Martin Rd.
Roswell, GA 30076

Dear Mr. Herbert:

In accordance with your request, the following are the proponents of Proposals 3 and 4 included in our 1990 Proxy Statement:

	Claim to Beneficially Own
Proposal #3	
Evangelical Lutheran Church in America 8765 West Higgins Road Chicago, IL	120,000 shares
Ed Crane, Director Corporate Social Responsibility	
Proposal #4 (co-sponsored) Adrian Dominican Sisters 1257 East Siena Heights Drive Adrian, MI	1,098 shares
Sister Annette M. Sinagra, O.P. Corporate Responsibility Coordinator	
and	
Corporate Responsibility Office Province of Saint Joseph of the Capachin Order 1534 Arch Street Berkeley, CA	40 shares
(Rev.) Michael H. Crosby, OFMCap Corporate Responsibility Agent	

Sincerely,

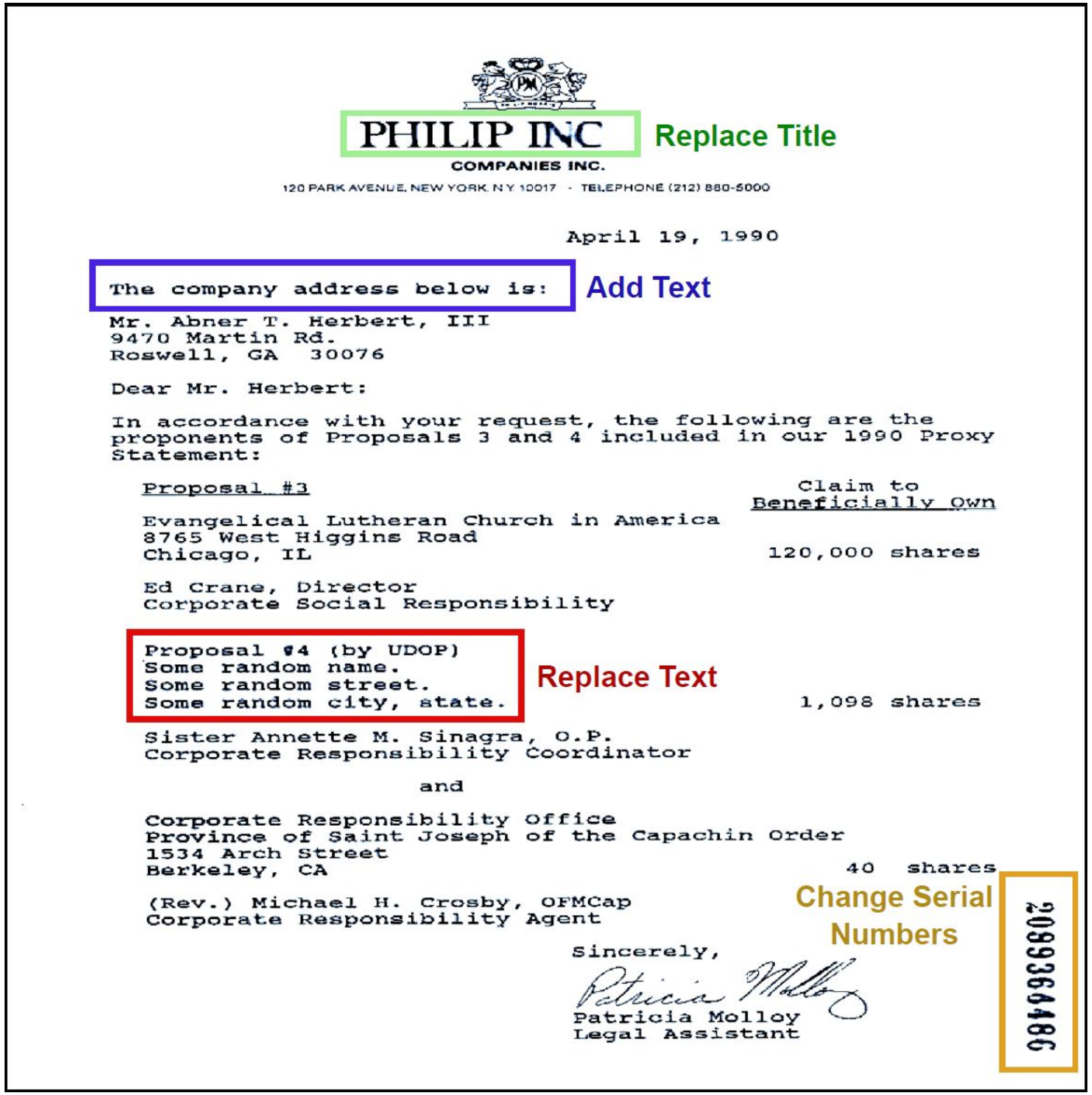
Patricia Molloy

Patricia Molloy
Legal Assistant

2048160205

Document image editing & customization with i-Code Doc

Edited Document With Customized Content



PHILIP INC

COMPANIES INC.

120 PARK AVENUE, NEW YORK, N.Y. 10017 - TELEPHONE (212) 880-5000

April 19, 1990

The company address below is:

Mr. Abner T. Herbert, III
9470 Martin Rd.
Roswell, GA 30076

Dear Mr. Herbert:

In accordance with your request, the following are the proponents of Proposals 3 and 4 included in our 1990 Proxy Statement:

Proposal #3	Claim to Beneficially Own
Evangelical Lutheran Church in America 8765 West Higgins Road Chicago, IL	120,000 shares
Ed Crane, Director Corporate Social Responsibility	
Proposal #4 (by UDOP) Some random name. Some random street. Some random city, state.	1,098 shares
Sister Annette M. Sinagra, O.P. Corporate Responsibility Coordinator	
and	
Corporate Responsibility Office Province of Saint Joseph of the Capachin Order 1534 Arch Street Berkeley, CA	40 shares
(Rev.) Michael H. Crosby, OFMCap Corporate Responsibility Agent	

Sincerely,

Patricia Molloy
Legal Assistant

2099364486

Document image editing & customization with i-Code Doc

Original Document

Mr. William J. Halley. Page Two.

According to scientific test results in our files as recent as two weeks ago, the smoke of Kent cigarettes does not contain significantly less nicotine or tar than the smoke from other filter tip cigarettes and, insofar as the claim "cleans" the smoke is concerned, contains substantial quantities of both nicotine and tars. The same test results indicate that the smoke from Kent cigarettes does not contain significantly less nicotine than another popular brand of non-filter cigarette and contains more tar in the smoke as compared to the smoke of the same non-filter popular brand cigarette. In view of this information, we question the propriety of the themes above referred to in the advertising for Kent cigarettes.

It is our view that smokers expect from advertising claims comparing the merits of filters and filtration effectiveness that they will benefit physically and to the degree claimed or implied in the advertising by the advertised brand.

We have further noted in connection with your advertising of Old Gold filter cigarettes the claim "one of the finest filters known to science." 4/ Our information indicates that five other filter brands contain less nicotine and tars in the smoke than Old Gold filter cigarette smoke contains. We therefore question this claim under Guide 2.

We are writing you in a spirit of cooperation to inquire what your company's diskposition and policy will be with respect to these and similar claims for the Kent and Old Gold cigarettes.

We will greatly appreciate a prompt response to this letter and your advice in this regard.

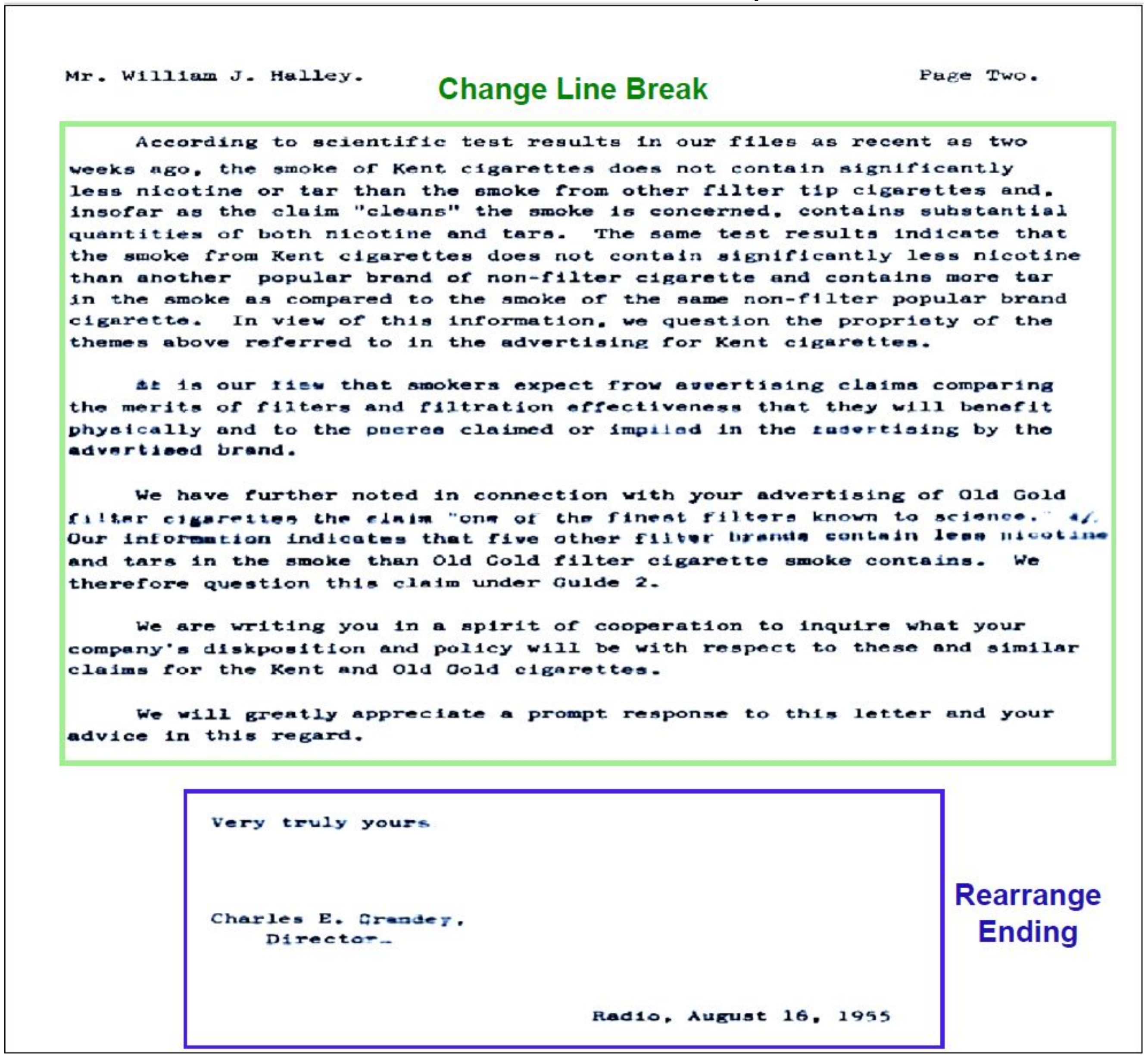
Very truly yours,

Charles E. Grandey,
Director.

4/ Radio, August 16, 1955

Document image editing & customization with i-Code Doc

Edited Document Layout



Mr. William J. Halley. Page Two.

According to scientific test results in our files as recent as two weeks ago, the smoke of Kent cigarettes does not contain significantly less nicotine or tar than the smoke from other filter tip cigarettes and, insofar as the claim "cleans" the smoke is concerned, contains substantial quantities of both nicotine and tars. The same test results indicate that the smoke from Kent cigarettes does not contain significantly less nicotine than another popular brand of non-filter cigarette and contains more tar in the smoke as compared to the smoke of the same non-filter popular brand cigarette. In view of this information, we question the propriety of the themes above referred to in the advertising for Kent cigarettes.

At is our view that smokers expect frow aveertising claims comparing the merits of filters and filtration effectiveness that they will benefit physically and to the pacree claimed or implied in the zadertising by the advertised brand.

We have further noted in connection with your advertising of Old Gold filter cigarettes the claim "one of the finest filters known to science." 4/. Our information indicates that five other filter brands contain less nicotine and tars in the smoke than Old Gold filter cigarette smoke contains. We therefore question this claim under Guide 2.

We are writing you in a spirit of cooperation to inquire what your company's diskposition and policy will be with respect to these and similar claims for the Kent and Old Gold cigarettes.

We will greatly appreciate a prompt response to this letter and your advice in this regard.

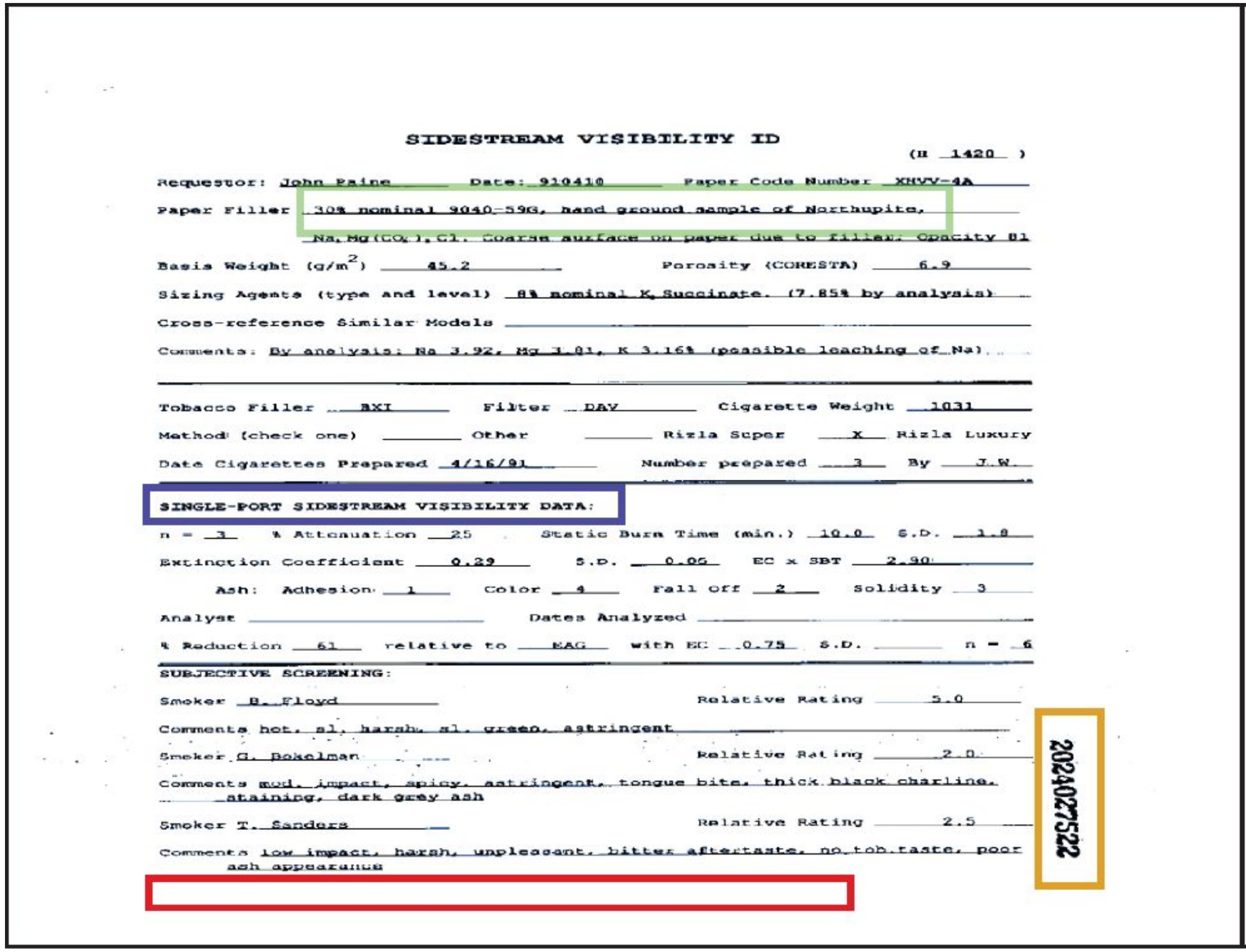
Very truly yours

Charles E. Grandey,
Director.

Radio, August 16, 1955

Document image editing & customization with i-Code Doc

Original Document



SIDESTREAM VISIBILITY ID

(H 1420)

Requestor: John Paine Date: 910410 Paper Code Number XHVV-4A

Paper Filler 30% nominal 9040-59G, hand ground sample of Northupite,
Na, Mg(CO_3), Cl. Coarse surface on paper due to filler. Opacity 81

Basis Weight (g/m^2) 45.2 Porosity (CORESTA) 6.9

Sizing Agents (type and level) 8% nominal K Succinate. (7.85% by analysis)

Cross-reference Similar Models

Comments: By analysis: Na 3.92, Mg 1.81, K 3.16% (possible leaching of Na).

Tobacco Filler BXI Filter DAV Cigarette Weight 1031

Method (check one) Other Rizla Super X Rizla Luxury

Date Cigarettes Prepared 4/16/91 Number prepared 3 By J.W.

SINGLE-PORT SIDESTREAM VISIBILITY DATA:

n = 3 % Attenuation 25 Static Burn Time (min.) 10.0 S.D. 1.8

Extinction Coefficient 0.29 S.D. 0.06 EC x SBT 2.90

Ash: Adhesion 1 Color 4 Fall off 2 Solidity 3

Analyst Dates Analyzed

% Reduction 61 relative to EAG with EC 0.75 S.D. n = 6

SUBJECTIVE SCREENING:

Smoker B. Floyd Relative Rating 5.0

Comments hot, sl. harsh, sl. green, astringent

Smoker G. Bokelman Relative Rating 2.0

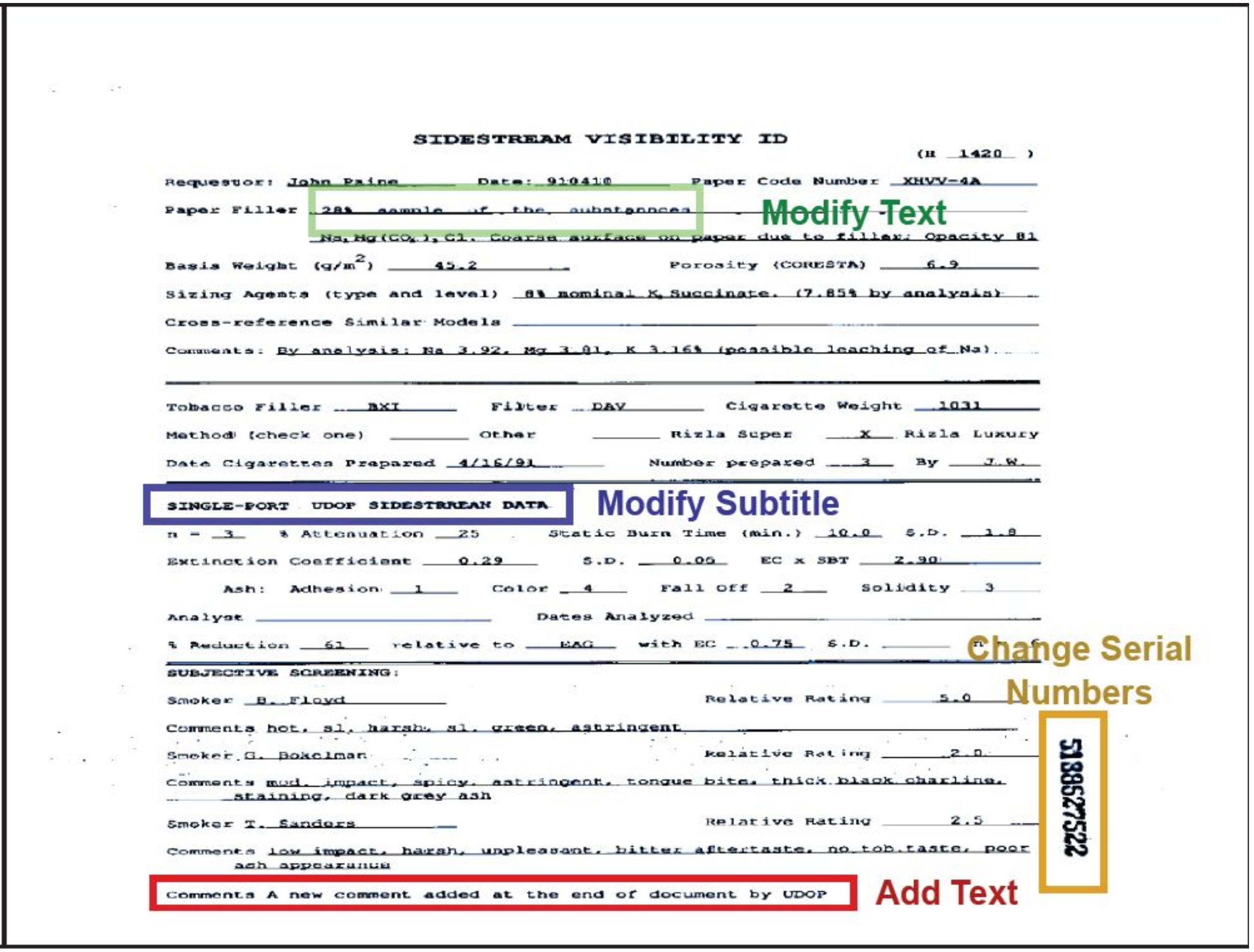
Comments mod. impact, spicy, astringent, tongue bite, thick black charline, staining, dark grey ash

Smoker T. Sanders Relative Rating 2.5

Comments low impact, harsh, unpleasant, bitter aftertaste, no tob.taste, poor ash appearance

2024027522

Edited Document



SIDESTREAM VISIBILITY ID

(H 1420)

Requestor: John Paine Date: 910410 Paper Code Number XHVV-4A

Paper Filler 28% sample of the substannces
Na, Mg(CO_3), Cl. Coarse surface on paper due to filler. Opacity 81

Basis Weight (g/m^2) 45.2 Porosity (CORESTA) 6.9

Sizing Agents (type and level) 8% nominal K Succinate. (7.85% by analysis)

Cross-reference Similar Models

Comments: By analysis: Na 3.92, Mg 1.81, K 3.16% (possible leaching of Na).

Tobacco Filler BXI Filter DAV Cigarette Weight 1031

Method (check one) Other Rizla Super X Rizla Luxury

Date Cigarettes Prepared 4/16/91 Number prepared 3 By J.W.

SINGLE-PORT UDOP SIDESTRREAN DATA

n = 3 % Attenuation 25 Static Burn Time (min.) 10.0 S.D. 1.8

Extinction Coefficient 0.29 S.D. 0.06 EC x SBT 2.90

Ash: Adhesion 1 Color 4 Fall off 2 Solidity 3

Analyst Dates Analyzed

% Reduction 61 relative to EAG with EC 0.75 S.D.

SUBJECTIVE SCREENING:

Smoker B. Floyd Relative Rating 5.0

Comments hot, sl. harsh, sl. green, astringent

Smoker G. Bokelman Relative Rating 2.0

Comments mod. impact, spicy, astringent, tongue bite, thick black charline, staining, dark grey ash

Smoker T. Sanders Relative Rating 2.5

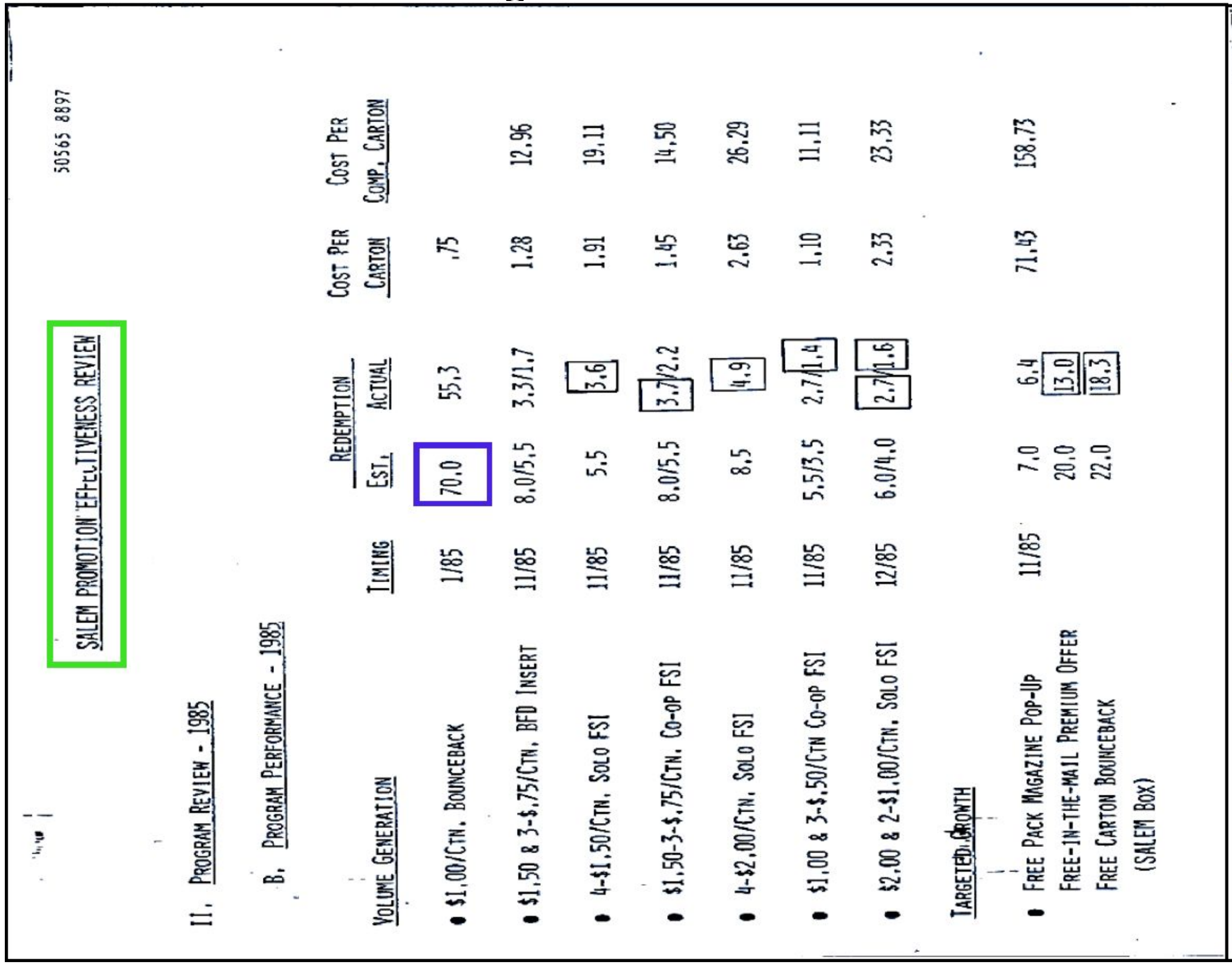
Comments low impact, harsh, unpleasant, bitter aftertaste, no tob.taste, poor ash appearance

Comments A new comment added at the end of document by UDOP

5188527522

Document image editing & customization with i-Code Doc

Original Document



50565 8897

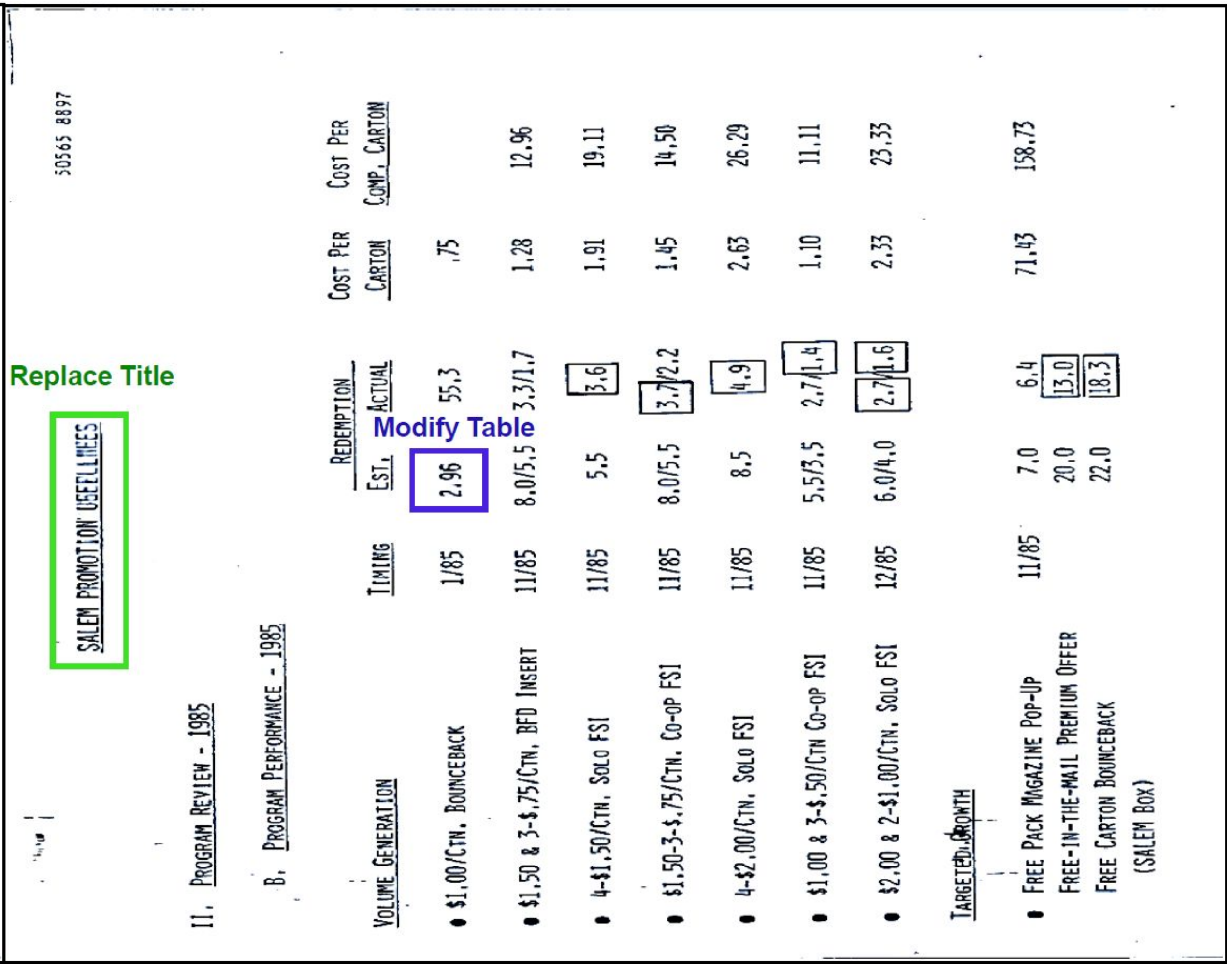
SALEM PROMOTION EFFECTIVENESS REVIEW

II. PROGRAM REVIEW - 1985

B. PROGRAM PERFORMANCE - 1985

	Timing	Redemption Est.	Redemption Actual	Cost Per Carton	Cost Per Comp. Carton
Volume Generation					
• $1.00/Ctn. Bounceback	1/85	70.0	55.3	.75	
• $1.50 & 3-$.75/Ctn. BFD Insert	11/85	8.0/5.5	3.3/1.7	1.28	12.96
• 4-$1.50/Ctn. Solo FSI	11/85	5.5	3.6	1.91	19.11
• $1.50-3-$.75/Ctn. Co-op FSI	11/85	8.0/5.5	3.7/2.2	1.45	14.50
• 4-$2.00/Ctn. Solo FSI	11/85	8.5	4.9	2.63	26.29
• $1.00 & 3-$.50/Ctn Co-op FSI	11/85	5.5/3.5	2.7/1.4	1.10	11.11
• $2.00 & 2-$1.00/Ctn. Solo FSI	12/85	6.0/4.0	2.7/1.6	2.33	23.33
Targeted Growth					
• Free Pack Magazine Pop-Up	11/85	7.0	6.4	71.43	158.73
Free-in-the-Mail Premium Offer		20.0	13.0		
Free Carton Bounceback (Salem Box)		22.0	18.3		

Edited Document



50565 8897

SALEM PROMOTION USEFLLNEES

II. PROGRAM REVIEW - 1985

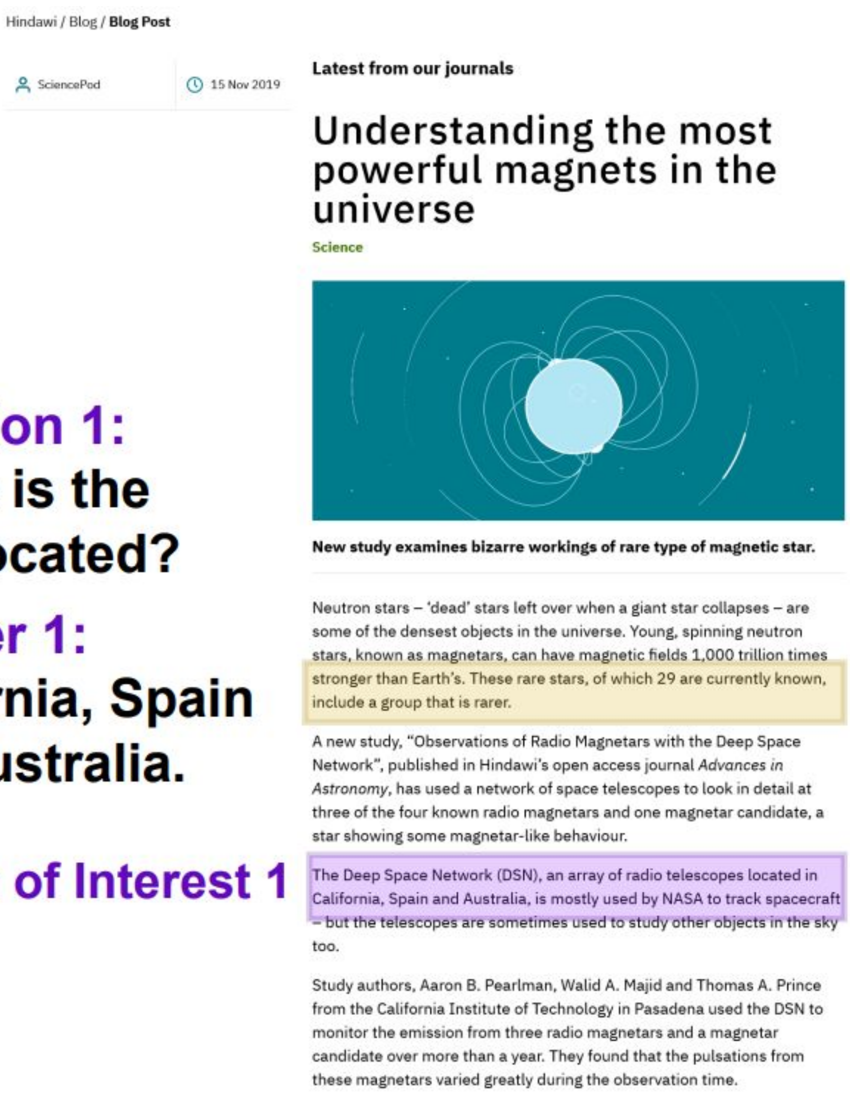
B. PROGRAM PERFORMANCE - 1985

	Timing	Redemption Est.	Redemption Actual	Cost Per Carton	Cost Per Comp. Carton
Volume Generation					
• $1.00/Ctn. Bounceback	1/85	2.96	55.3	.75	
• $1.50 & 3-$.75/Ctn. BFD Insert	11/85	8.0/5.5	3.3/1.7	1.28	12.96
• 4-$1.50/Ctn. Solo FSI	11/85	5.5	3.6	1.91	19.11
• $1.50-3-$.75/Ctn. Co-op FSI	11/85	8.0/5.5	3.7/2.2	1.45	14.50
• 4-$2.00/Ctn. Solo FSI	11/85	8.5	4.9	2.63	26.29
• $1.00 & 3-$.50/Ctn Co-op FSI	11/85	5.5/3.5	2.7/1.4	1.10	11.11
• $2.00 & 2-$1.00/Ctn. Solo FSI	12/85	6.0/4.0	2.7/1.6	2.33	23.33
Targeted Growth					
• Free Pack Magazine Pop-Up	11/85	7.0	6.4	71.43	158.73
Free-in-the-Mail Premium Offer		20.0	13.0		
Free Carton Bounceback (Salem Box)		22.0	18.3		

Agenda

- Document AI Background, Challenges & Motivations
- Model Architecture
- Pretraining
- Evaluations
- Controllable Document Image Generation
- Analysis

Answer Localization for Document QA



Hindawi / Blog / **Blog Post**

ScienccePod | 15 Nov 2019

Latest from our journals

Understanding the most powerful magnets in the universe

Science

New study examines bizarre workings of rare type of magnetic star.

Neutron stars – 'dead' stars left over when a giant star collapses – are some of the densest objects in the universe. Young, spinning neutron stars, known as magnetars, can have magnetic fields 1,000 trillion times stronger than Earth's. These rare stars, of which 29 are currently known, include a group that is rarer.

A new study, "Observations of Radio Magnetars with the Deep Space Network", published in Hindawi's open access journal *Advances in Astronomy*, has used a network of space telescopes to look in detail at three of the four known radio magnetars and one magnetar candidate, a star showing some magnetar-like behaviour.

The Deep Space Network (DSN), an array of radio telescopes located in California, Spain and Australia, is mostly used by NASA to track spacecraft – but the telescopes are sometimes used to study other objects in the sky too.

Study authors, Aaron B. Pearlman, Walid A. Majid and Thomas A. Prince from the California Institute of Technology in Pasadena used the DSN to monitor the emission from three radio magnetars and a magnetar candidate over more than a year. They found that the pulsations from these magnetars varied greatly during the observation time.

Share Post

Question 1:
Where is the DSN located?

Answer 1:
California, Spain and Australia.

Region of Interest 1

Question 2:
How many magnetars are known to people?

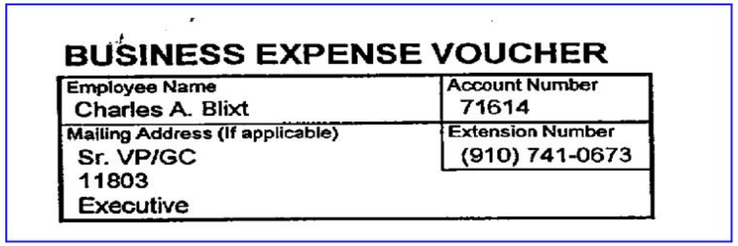
Answer 2:
29

Region of Interest 2

Note that VisualMRC dataset (the shown example) only provides paragraph-level answer locations

How effective is the vision modality?

DocVQA


BUSINESS EXPENSE VOUCHER

Employee Name Charles A. Blixt	Account Number 71614
Mailing Address (If applicable) Sr. VP/GC 11803 Executive	Extension Number (910) 741-0673

Q: What is the Extension Number as per the voucher?
GT: (910) 741-0673

The answer can be found in OCR annotations

Infographic VQA (InfoVQA)

How many companies have more than 10K delivery workers?
Answer: 2 Evidence: Figure
Answer-source: Non-extractive Operation: Counting Sorting

Who has better coverage in Toronto - Canada post or Amazon?
Answer: canada post Evidence: Text
Answer-source: Question-span Image-span Operation: *none*

In which cities did Canada Post get maximum media coverage?
Answer: vancouver, montreal Evidence: Text Map
Answer-source: Multi-span Operation: *none*

Figure 1: Example image from InfographicVQA along with questions and answers. For each question, source of the answer, type of evidence the answer is grounded on, and the discrete operation required to find the answer are shown.

Model	DocVQA	InfoVQA
UDOP	**84.7**	**47.4**
UDOP w/o Image Embeds	84.4	45.0

Yes, the vision modality is effective on **vision-dependent** dataset

Ablation Study

Self supervised pretraining is already **competitive**:

Table 8. Ablation study on pretraining objectives. Performance is reported on validation sets.

Pretrain Objectives	#Pretrain Data	DocVQA	RVL-CDIP	FUNSD	CORD
MLM	11.0M	79.7	95.3	90.2	96.7
UDOP-Dual					
Self-Supervised	11.0M	83.5	95.8	91.5	97.2
+ Supervised	12.8M	**84.1**	**96.1**	**91.5**	**97.3**
UDOP					
Self-Supervised	11.0M	84.4	96.2	91.0	97.2
+ Supervised	12.8M	**85.0**	**96.3**	**91.9**	**97.4**

Model	Modality	Info Ext.		Classification
		FUNSD	CORD	RVL-CDIP
Donut	V	-	91.6	95.3
$BERT_{large}$	T	65.63	90.25	89.92
$BROS_{large}$ [15]	T+L	84.52	97.40	-
$StructuralLM_{large}$	T+L	85.14	-	96.08
LiLT [47]	T+L	88.41	96.07	95.68
FormNet [23]	T+L	84.69	97.28	-
$LayoutLM_{large}$	T+L	77.89	-	91.90
SelfDoc	V+T+L	83.36	-	92.81
UDoc	V+T+L	87.93	98.94	95.05
$DocFormer_{large}$ [1]	V+T+L	84.55	96.99	95.50
$TILT_{large}$	V+T+L	-	96.33	95.52
$LayoutLMv2_{large}$	V+T+L	84.20	96.01	95.64
$LayoutLMv3_{large}$	V+T+L	**92.08**	97.46	95.93
UDOP-Dual	V+T+L	91.20	**97.64**	**96.22**
UDOP	V+T+L	**91.62**	**97.58**	**96.00**

Conclusion

- Unified **representations** and **modeling** for **vision, text and layout** modalities in document AI.
- Unified all document tasks to the **seq2seq generation** framework.
- Combined **novel self-supervised** objectives with **supervised datasets** in pretraining for unified document pretraining.
- UDOP can process and **generate text, vision, and layout** modalities together, which to the best of our knowledge is first one in the field of document AI.