

Visually-Rich Documentを軸 とした多言語処理の動向

2023年12月12日@名古屋

2024年01月12日@仙台

2024年7月23日@東京

藤沼祥成

自己紹介 : 略歴



International School of Paris
Educating for complexity

Amazon JP
2014 - 2016



東京大学
THE UNIVERSITY OF TOKYO



大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

2012 - 2014



Artes et Scientiae

2008 - 2012

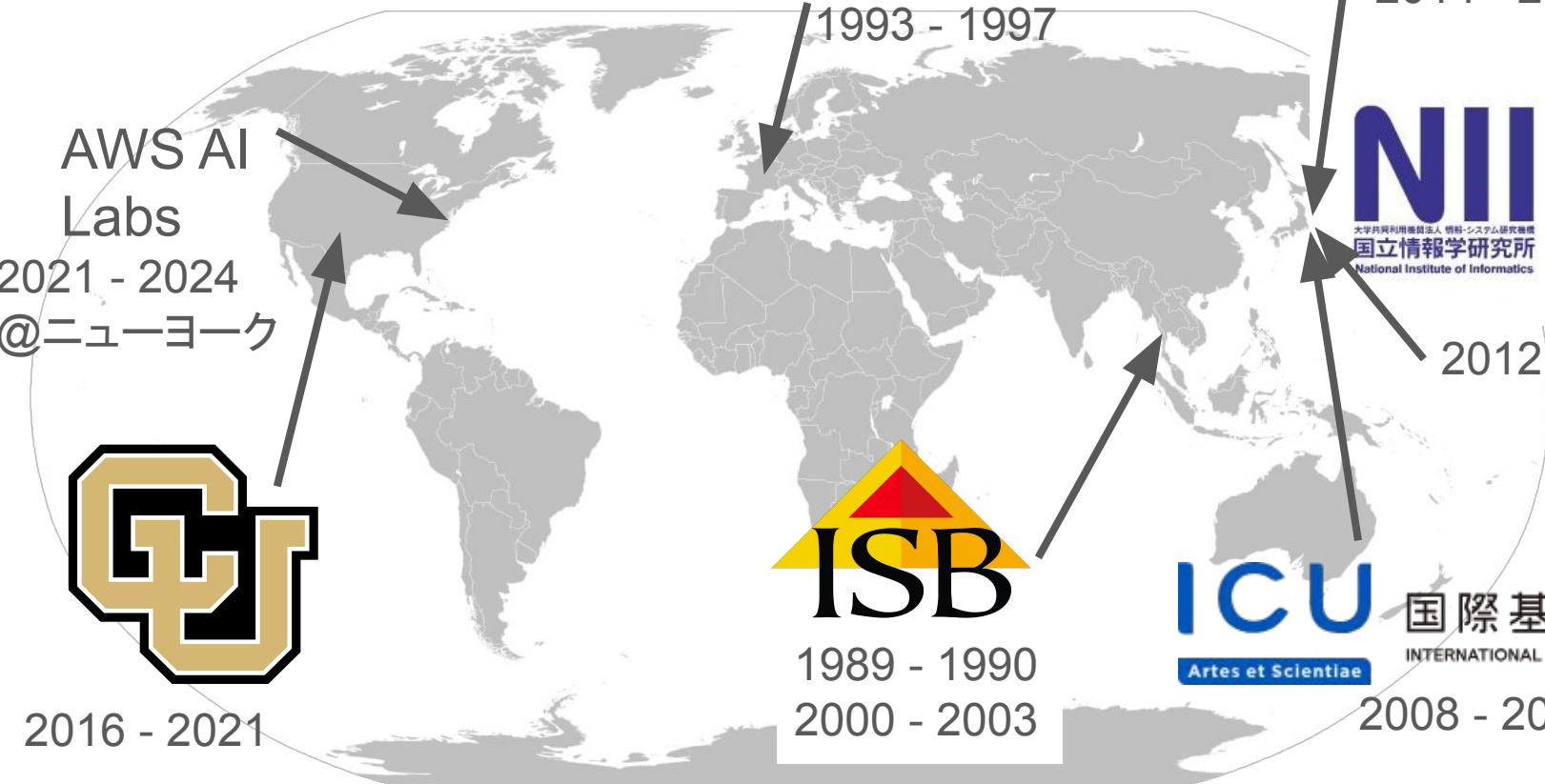
1989 - 1990
2000 - 2003



AWS AI
Labs
2021 - 2024
@ニューヨーク



2016 - 2021



自己紹介: 最近やっていること

- タイの出生証明書→→→→→→→→→→→→
- 2023年上半年はPDF/文書処理に関する仕事を主に従事
 - 多言語データセット構築の話がEMNLP Findings 2023に

去年の仕事だが今回話さないこと

- Dialogue State Trackingの効率化 [Lesci+ 2023]
- 多言語LMのバイアス解析 [Levy+ 2023]



タイの出生証明書例。在インドタイ大使館のサイトより引用

目次

- イントロ: 多言語NLP
- Visually-Rich Documentの多言語NLP
- Visually-Rich Documentから見たVision-Languageモデル

イントロ: 多言語 NLP

世界で話されている言語は7000言語以上



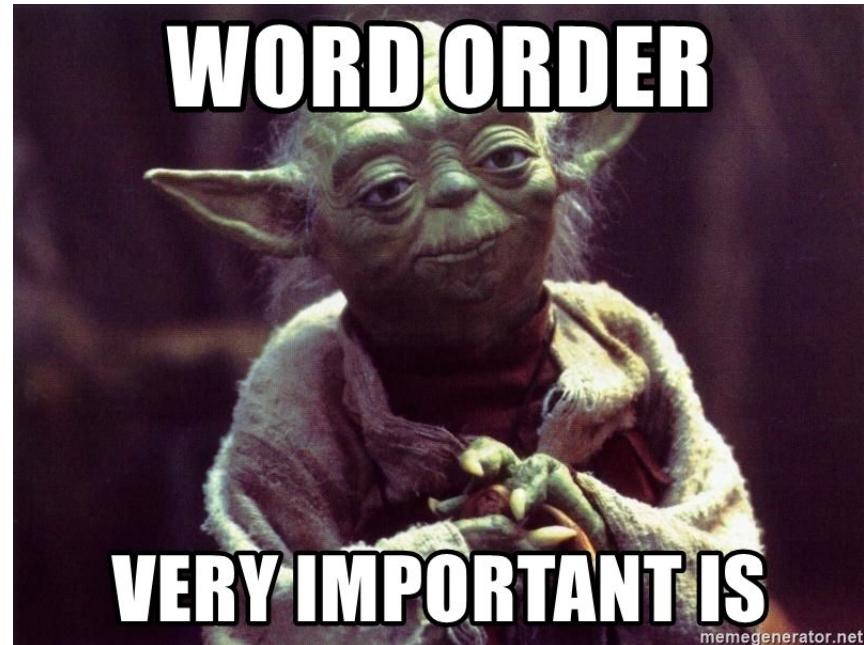
Ref: https://www.eventplanner.net/news/8886_how-to-run-an-efficient-multilingual-conference.html

多言語NLPにおける課題

- 例えはあるモデルは英語で学習して評価済み
- 他の言語では?
 - スペイン語は動くかも
 - アムハラ語では?

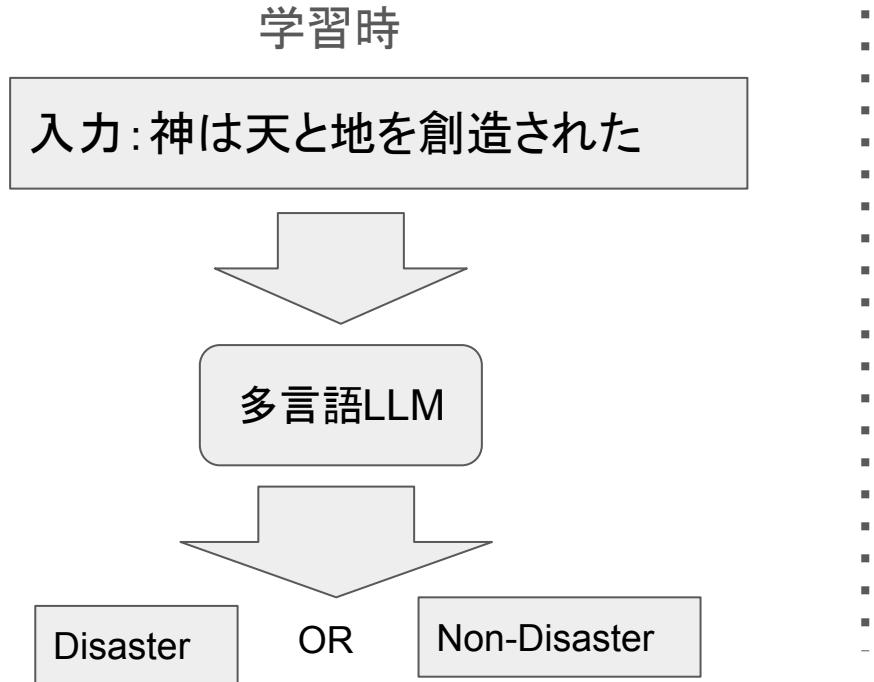
課題:

- 言語学観点から
 - 言語の多様性
- 機械学習的観点から
 - 対象言語のデータがない



言語間転移学習 (Cross-Lingual Transfer Learning)

- 言語1で学習し、言語2で推論する

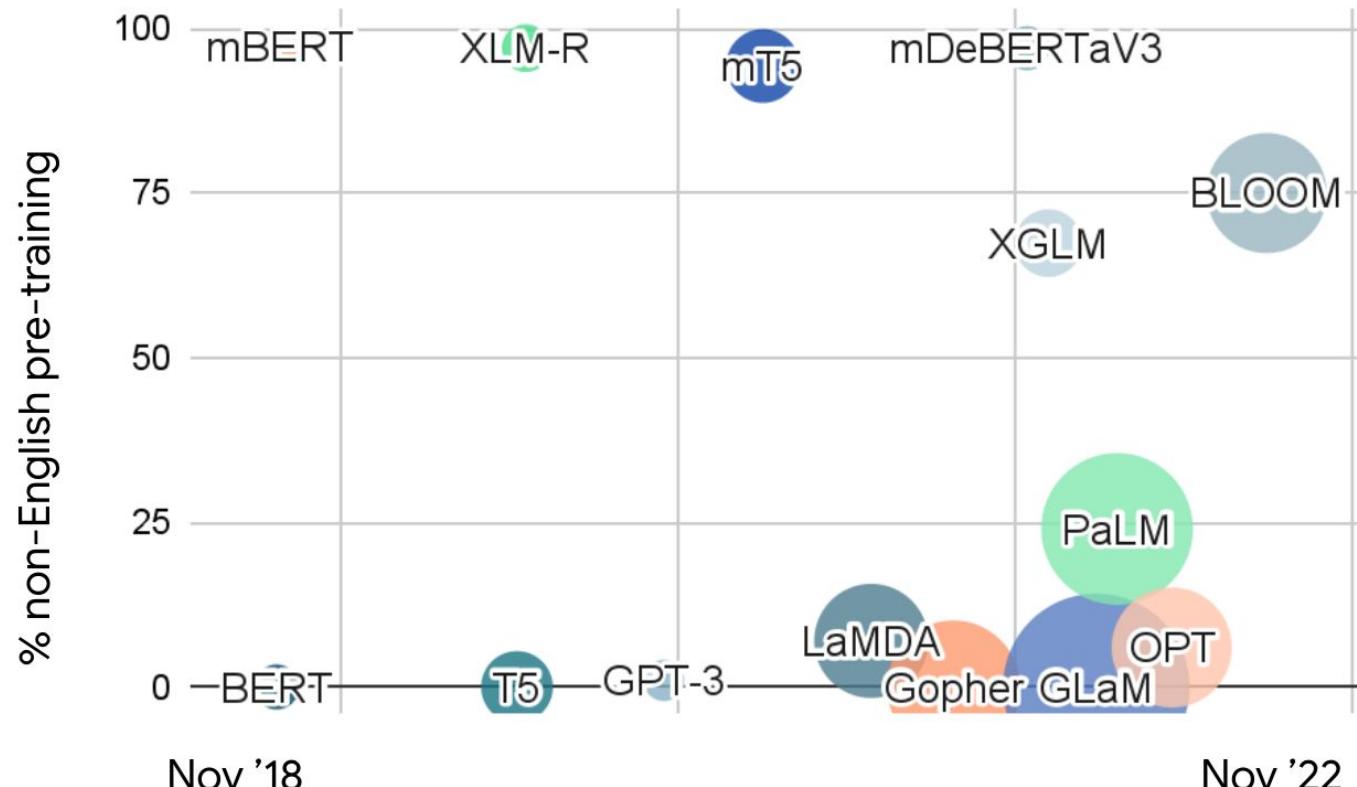


言語間転移学習 (Cross-Lingual Transfer Learning)

- 言語1で学習し、言語2で推論する

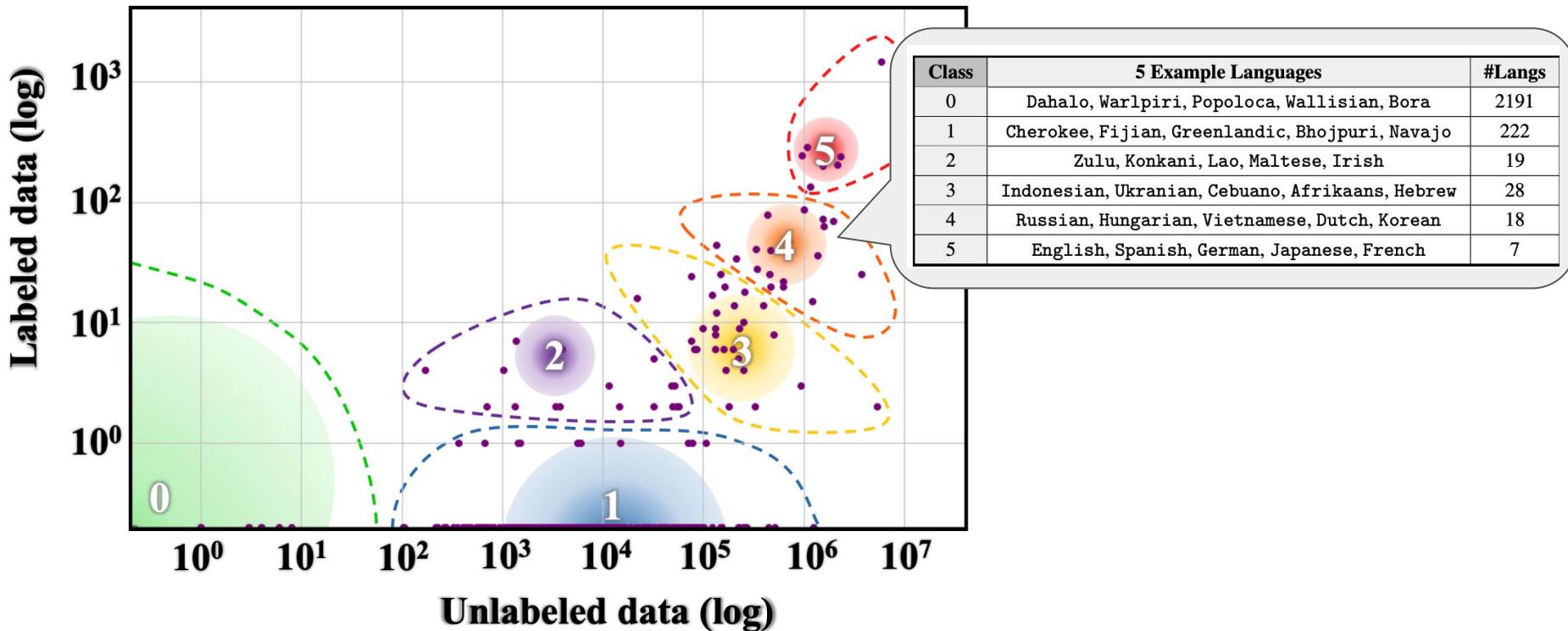


多言語事前学習済みモデルは多くある



Ref: <https://www.ruder.io/state-of-multilingual-ai/>

各言語のデータのボトルネックをどうするかが今後の課題



Visually-Rich Documentの多言語研究

Visually-Rich Documentとは何か?

ウィキペディア

三 ...

出典: フリー百科事典「ウィキペディア (Wikipedia)」

カリビアンカップ2010

ページ ノート その他 出典: フリー百科事典「ウィキペディア (Wikipedia)」

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。

予選 [編集] 詳細は「カリビアンカップ2010予選」を参照

出場国 [編集]

- マルティニーク (開催国)
- ジャマイカ (開催国)
- キューバ (予選突破)
- グレナダ (予選突破)
- トリニダード・トバゴ (予選突破)
- グアドループ (予選突破)
- アンティグア・バーブーダ (予選突破)
- ガイアナ (予選突破)

グループステージ [編集] 時間は全て (UTC-4)。

グループH [編集]

Team	P	D	W	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニーク	3	0	1	2	1	3	-2	1

Tables

Images

2010年11月26日 18:00 トリニダード・トバゴ 0-2 Report

スタッド・ビエール=アキケ、フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランクスター (ガイアナ)

Eurlex: EUの法令関連文書

16.3.2006 EN Official Journal of the European Union L 79/27

COMMISSION

COMMISSION DECISION

of 6 March 2006

establishing the classes of reaction-to-fire performance for certain construction products as regards

wood flooring and solid wood paneling and cladding

(notified under document number C(2006) 655)

(Text with EEA relevance)

(2006/213/EC)

THE COMMISSION OF THE EUROPEAN COMMUNITIES.

Having regard to the Treaty establishing the European Community,

Having regard to Directive 89/106/EEC of 21 December 1988, on the approximation of laws, regulations and administrative provisions of the Member States relating to construction products (1), and in particular Article 20(2) thereof;

Whereas:

(1) Directive 89/106/EEC envisions that in order to take account of different levels of protection for construction works at national, regional or local level, it may be necessary to establish in the interpretative documents clauses defining the reaction-to-fire performance of products in respect of each essential requirement. Those documents have been published as the 'Communication of the Commission with regard to the interpretative documents of Directive 89/106/EEC' (2).

(2) With respect to the essential requirement of safety in the event of fire, interpretative document No 2 lists a number of interrelated measures which together define the fire safety strategy to be variously developed in the Member States.

(3) Interpretative document No 2 identifies one of those measures as the limitation of the generation and spread of fire and smoke within a given area by limiting the potential of construction products to contribute to the full development of a fire.

(4) The level of that limitation may be expressed only in terms of the different levels of reaction-to-fire performance of the products in their end-use application:

(1) OJ L 40, 11.2.1989, p. 12. Directive as last amended by Regulation (EC) No 1882/2003 of the European Parliament and the Council (OJ L 284, 29.10.2003, p. 1).

(2) OJ C 1, 28.2.1994, p. 1.

(3) OJ L 59, 23.3.2000, p. 14. Decision as amended by Decision 2001/132/EC (OJ L 220, 3.9.2001, p. 5).

(5) By way of harmonised solution, a system of classes was adopted in Commission Decision 2000/147/EC of 8 February 2000 implementing Council Directive 89/106/EEC as regards the classification of the reaction-to-fire performance of construction products (3).

(6) In the case of wood flooring and solid wood paneling and cladding it is necessary to use the classification established in Decision 2000/147/EC.

(7) The reaction-to-fire performance of many construction products and materials within the classification provided for in Decision 2000/147/EC is well established and sufficiently well known to fire regulators in Member States that they do not require testing for this particular performance characteristic.

(8) The measures provided for in this Decision are in accordance with the opinion of the Standing Committee on Construction.

HAS ADOPTED THIS DECISION:

Article 1

The construction products and/or materials which satisfy all the requirements of the performance characteristic 'reaction to fire' without need for further testing are set out in the Annex.

Article 2

The specific classes to be applied to different construction products and/or materials, within the reaction-to-fire classification adopted in Decision 2000/147/EC, are set out in the Annex to this Decision.

(1) OJ L 59, 23.3.2000, p. 14. Decision as amended by Decision 2001/132/EC (OJ L 220, 3.9.2001, p. 5).

Visually-Rich Documentを処理する上での課題

ウィキペディア

カリビアンカップ2010

ページ ノート ソリューション 出典: フリー百科事典 「Wikivisual」 (Wikivisual)

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFサッカーワールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。

予選 [編集]

詳細版「[カリビアンカップ2010予選](#)」を参照

出場国 [編集]

- マルティニーク (開催国)
- ジャマイカ (前回優勝)
- キューバ (予選突破)
- グレナダ (予選突破)
- トリニダード・トバゴ (予選突破)
- グアドループ (予選突破)
- アンティグア・バーブーダ (予選突破)
- ガイアナ (予選突破)

グルーブステージ [編集]

時間は全て (UTC-4)。

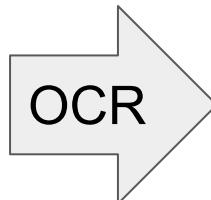
グループH [編集]

Team	Pld	W	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニーク	3	0	1	2	1	3	-2	1

2010年11月26日
18:00
トリニダード・トバゴ  0-2 Report

スタッド・ピエール=アキエ、フォールド・ド・フランス
観客数: 5,000
主審: ハビエル・エスコベド (メキシコ)
VAR: デイビッド・マーティン (メキシコ)
アシスタント: ホセ・アントニオ・モラレス (メキシコ)
4G: 339
リナーグ: 99

Tables
Images



Table

Images



ウィキペディア フリー

[...]

グループステージ[編集]時間は全て(UTC-4)グループH[編集]グループH[編集]Team Plid WDL GF GA GD Pts キューバ 3 2 i 0 3 0 +3 7

[...]

Visually-Rich Documentを処理する上での課題

ウィキペディア

三 ウィキペディア フリー百科事典

...
カリビアンカップ2010

ページ ノート その他 出典: フリー百科事典「ウィキペディア (Wikipedia)」

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニクで開催された。

予選 [編集]
詳細は「[カリビアンカップ2010予選](#)」を参照

出場国 [編集]
• マルティニク (開催国)
• ジャマイカ (前回優勝)
• キューバ (予選突破)
• グレナダ (予選突破)
• トリニダード・トバゴ (予選突破)
• グアドループ (予選突破)
• アンティグア・バーブーダ (予選突破)
• ガイアナ (予選突破)

グループステージ [編集]
時間は全て (UTC-4)。
グループH [編集]

Team	P	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3 7
グレナダ	3	1	2	0	2	1	+1 5
トリニダード・トバゴ	3	1	0	2	1	3	-2 3
マルティニク	3	0	1	2	1	3	-2 1

2010年11月26日
18:00
トリニダード・トバゴ 0-2 Report
0-2
スタッド・ビエール=アキケー, フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

J. コロナ 339
リナレ 99

イニシアチブ

OCR

Tables
Images

画像はOCR
で出力なし

ウィキペディア フリー

[...]

グループステージ[編集]時間は全て(UTC-4)グループH[編集]グループH[編集]Team Pld W D L GF GA GD Pts キューバ 3 2 1 0 3 2 0 +3 7

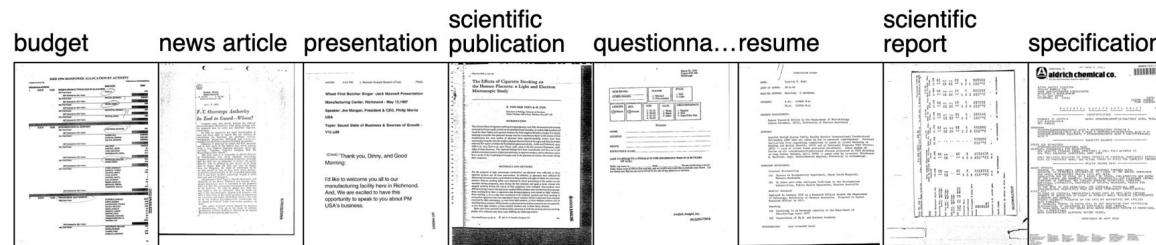
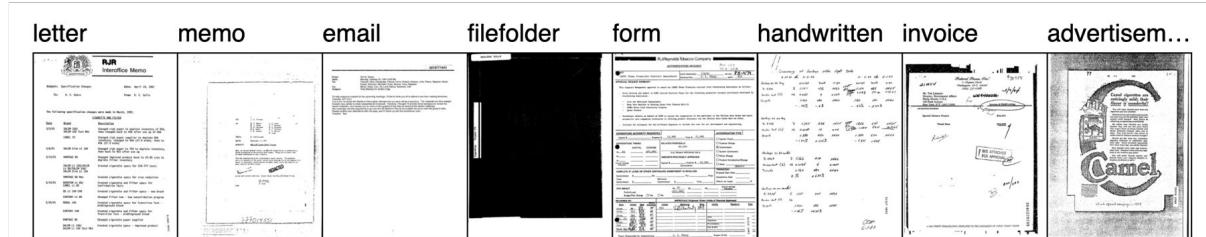
0 +3 7

[...]

OCRエラー

Visually-Rich Documentの分類データセット[Harley+ 2015]

- 文書画像の16種類の文書タイプ(e.g., メール vs. メモ)に分類
- 問題点
 - 文書の中身を理解しなくても分類可能
 - 英語のみ



英語以外のVisually-Rich Documentは？

- 8言語の文書画像理解データセットは存在する [Xu+ 2021, Wang+ 2022]
- 他の言語、他のタスクにおいてモデルをどう評価するか？

Submitting information					
*Sender	Xinran Rong				
*Address	Ytong Manchu Autonomous County, Spring City, Jin Province				
*Telephone number	13030257941				
*Destination country	China				
*Method of send report	<input checked="" type="checkbox"/> E-mail <input type="checkbox"/> Postal <input type="checkbox"/> Sender <input type="checkbox"/> Owner <input type="checkbox"/> By yourself				
*Owner's name	Ying Chong				
*Address	Nanping County, Zhangzhou City, Fujian Province				
*Telephone number	13148397260				
Animal's details					
*Microchip number	89512496411251264				
*Species	Dog	Breed	Shiba Inu		
Name	Marmo	Age	3 months.	Sex	<input type="checkbox"/> Male [♂] <input checked="" type="checkbox"/> Female [♀]
Date of last rabies vaccination	2019.9.5				
Date of sampling	2020.3.21				
Blood collection department	Strength pet hospital				
Blood sample collector	Jiang Li				
Vaccine make	China Biotechnology Corporation				
*Disclaimer: The information provided therein above is true and valid. If any falsify, I will take all the compensation and legal liability caused.					
Signature Xinran Rong					
Reference laboratory use only					
Date	2020.4.26	Sample No	8962146521415		
实验室使用参考用					
Date	2020.4.26	样品编号	8962146521415		

(a) A form.

Image from [Wang+ 2022]

HAPPY HOUR RESTAURANT

Invoice # : 1803321345698 TEL: 07-3233123

Table : 3

Member Points : 0.00

Date: 20/01/2019 Time: 12:42:00

Cashier: He Bai

Item	Qty	U/P	DISC %	Amount
Seafood Noodles	238	1	10.00	10.00
Milk	1033	1	3.00	3.00
Total Qty :	2			
Total Points :	0.00			

Sub Total : 13.00
 DISC : 0.00
 Service Charge : 0.00
 Tax : 0.80

Total	13.80
Cash	13.80
Change	0.00

Goods sold are not returnable.
 Thank you !

欢乐时光餐厅

发票编号: 1803321345698 电话: 07-3233123

桌号: 3

会员积分: 0.00

日期: 20/01/2019 时间: 12:42:00

收银员: 百合

品名	数量	单价	折扣 %	金额
海鲜面条	238	1	10.00	10.00
牛奶	1033	1	3.00	3.00
总计件数:	2			
总积分:	0.00			

小计: 13.00
 折扣: 0.00
 服务费: 0.00
 合税: 0.80

总计 13.80
现金 13.80
找零 0.00

商品一经售出概不退还。
 谢谢！

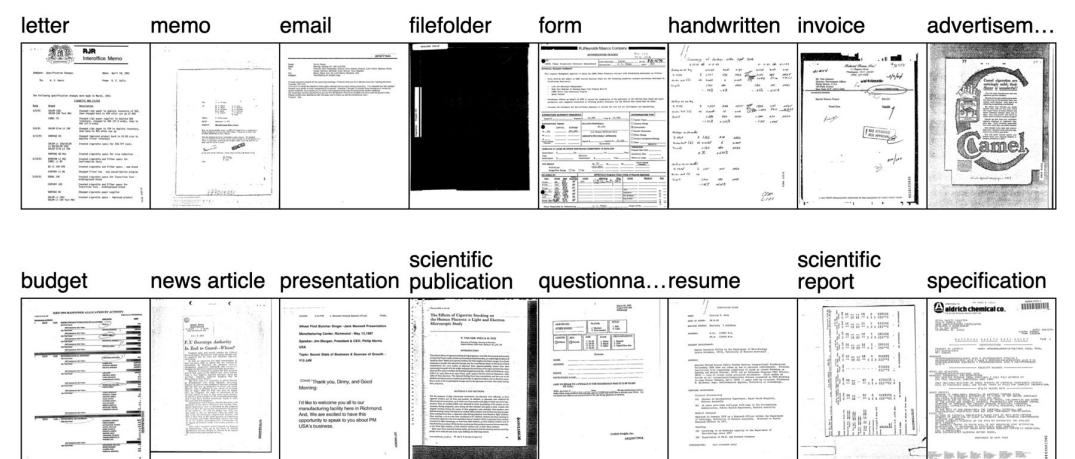
(b) A receipt.

Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータの良い点:

- トピック分類タスクなので文書内容が重要

リマインド: RVL-CDIP [Harley+ 2015]: 16
文書タイプラベル が付与された 文書画像
分類データセット



Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータの良い点:

- トピック分類タスクなので文書内容が重要

Doc Type: Internet Website

Tables

Images

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。

予選 [編集]
詳細は「[カリビアンカップ2010予選](#)」を参照

出場国 [編集]
● マルティニーク (開催国)
● ジャマイカ (前回優勝)
● ドミニカ (予選突破)
● グレナダ (予選突破)
● トリニダード・トバゴ (予選突破)
● フランス (予選突破)
● アンティグア・バーブーダ (予選突破)
● カリブ海 (予選突破)

グループステージ [編集]
時間は全て (UTC-4).

Team	Pld	W	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニーク	3	0	1	2	1	3	-2	1

2010年11月26日 18:00
トロニダード・トバゴ 0-2 Report
スタッド・ビエール=アキエ, フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

● キューバ
● コロ
● リナレ
● フラ
● フラ
● フラ
● フラ

Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータの良い点:

- トピック分類タスクなので文書内容が重要

...
Wikibedia
フリー百科事典
☰
カリビアンカップ2010

ページ ノート その他 出典: フリー百科事典「Wikibedia (Wikipedia)」

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。

予選 [編集]
詳細は「カリビアンカップ2010予選」を参照

出場国 [編集]
• **マルティニーク** (開催国)
• **ジャマイカ** (前回優勝)
• **キューバ** (予選突破)
• **グレナダ** (予選突破)
• **トリニダード・トバゴ** (予選突破)
• **グアドループ** (予選突破)
• **アンティグア・バーブーダ** (予選突破)
• **ガイアナ** (予選突破)

グループステージ [編集]
時間は全て (UTC-4).

グループH [編集]

Team	Pld	W	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニーク	3	0	1	2	1	3	-2	1

2010年11月26日
18:00
トロント・エリザベス・スタジアム
トロント
0-2
Report

スタッド・ビール=アキーエ, フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

Doc Topic:
SoccerTournament

Tables

Images

Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータの良い点:

- トピック分類タスクなので文書内容が重要
- マルチラベル(1文書に1つ以上のラベル)

16.3.2006 | EN | Official Journal of the European Union | L 79/27

COMMISSION

COMMISSION DECISION
of 6 March 2006
establishing the classes of reaction-to-fire performance for certain construction products as regards
wood flooring and solid wood paneling and cladding
(notified under document number C(2006) 655)
(Text with EEA relevance)
(2006/213/EC)

THE COMMISSION OF THE EUROPEAN COMMUNITIES,
Having regard to the Treaty establishing the European Community,
Having regard to Directive 89/106/EEC of 21 December 1988,
on the approximation of laws, regulations and administrative provisions of the Member States relating to construction products (1), and in particular Article 2(2) thereof,
Whereas:
(1) Directive 89/106/EEC envisages that in order to take account of different levels of protection for construction works at national, regional or local level, it may be necessary to establish different reaction-to-fire performance classes corresponding to the performance of products in respect of each essential requirement. Those documents have been published as the Communication of the Commission with regard to the interpretative documents of Directive 89/106/EEC (2).
(2) With respect to the essential requirement of safety in the event of fire, interpretative document No 2 lists a number of interrelated measures which together define the fire safety strategy to be variously developed in the Member States.
(3) Interpretative document No 2 identifies one of those measures as being the limitation of the spread of fire and smoke within a given area by means of the potential of construction products to contribute to the full development of a fire.
(4) The level of that limitation may be expressed only in terms of the different levels of reaction-to-fire performance of the products in their end-use application:
(i) OJ L 40, 11.2.1989, p. 12. Directive as last amended by Regulation (EC) No 1852/2004 of the European Parliament and the Council (2) OJ L 220, 28.7.2004, p. 1.
(3) OJ C 62, 28.2.1994, p. 1.

Label 1: Industry

Label 2: Environment

Label 3: Production

Label 4: Trade

Label 5: Social questions

Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータの良い点:

- トピック分類タスクなので文書内容が重要
- マルチラベル(1文書に1つ以上のラベル)
- 多言語
 - EU言語+アラビア語+中国語+日本語

16.3.2006 | EN | Official Journal of the European Union | L 79/27

COMMISSION
COMMISSION DECISION
of 6 March 2006
establishing the classes of reaction-to-fire performance for certain construction products as regards
wood flooring and solid wood paneling and cladding
(notified under document number C(2006) 655)
(Text with EEA relevance)
(2006/213/EC)

THE COMMISSION OF THE EUROPEAN COMMUNITIES,

Having regard to the Treaty establishing the European Community;

Having regard to Directive 89/106/EEC of 21 December 1988, on the approximation of laws, regulations and administrative provisions of the Member States relating to construction products (1), and in particular Article 20(2) thereof;

Whereas:

(1) Directive 89/106/EEC envisages that in order to take account of different levels of protection for construction works at national, regional or local level, it may be necessary to establish different reaction-to-fire performance classes corresponding to the performance of products in respect of each essential requirement. Those documents have been published as the Communication of the Commission with regard to the interpretative documents of Directive 89/106/EEC (2).

(2) With respect to the essential requirement of safety in the event of fire, interpretative document No 2 lists a number of interrelated measures which together define the fire safety strategy to be variously developed in the Member States.

(3) Interpretative document No 2 identifies one of those measures as being the reaction-to-fire and spread of fire and smoke within a given area by the use of the potential of construction products to contribute to the full development of a fire.

(4) The level of that limitation may be expressed only in terms of the different levels of reaction-to-fire performance of the products in their end-use application:

(i) OJ L 40, 11.2.1989, p. 12. Directive as last amended by Regulation (EC) No 1882/2002 of the European Parliament and the Council (OJ L 281, 29.10.2002, p. 1);

(ii) OJ C 62, 28.2.1994, p. 1.

COMMISSION

COMMISSION DECISION

of 6 March 2006

establishing the classes of reaction-to-fire performance for certain construction products as regards

wood flooring and solid wood paneling and cladding

(notified under document number C(2006) 655)

(Text with EEA relevance)

(2006/213/EC)

(5) By way of harmonised solution, a system of classes was adopted in Commission Decision 2000/147/EC of 8 February 2000 implementing Council Directive 89/106/EEC as regards the classification of the reaction-to-fire performance of construction products (3).

(6) In the case of wood flooring and solid wood paneling and cladding it is necessary to use the classification established in Decision 2000/147/EC.

(7) The reaction-to-fire performance of many construction products and/or materials within the meaning of Annex I to Directive 2000/147/EC is well established and sufficiently well known to fire regulators in Member States that they do not require testing for this particular performance characteristic.

(8) The measures provided for in this Decision are in accordance with the opinion of the Standing Committee on Construction.

HAS ADOPTED THIS DECISION:

Article 1

The construction products and/or materials which satisfy all the requirements of the performance characteristic 'reaction to fire' without need for further testing are set out in the Annex.

Article 2

The specific classes to be applied to different construction products and/or materials within the reaction-to-fire classification adopted in Decision 2000/147/EC, are set out in the Annex to this Decision.

(i) OJ L 50, 23.2.2000, p. 14. Decision as amended by Decision 2003/632/EC (OJ L 226, 3.9.2003, p. 9).

≡ ...
维基百科
维基百科是...
≡
カリビアンカップ2010
ページ ノート その他 出典 フリー百科事典「维基百科」(Wikipedia).
カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。
予選 (国際)
詳細は「カリビアンカップ2010予選」を参照
出場国 (国際)
• 法国 (マルティニーク (連邦国))
• 委内瑞拉 (前国民委員会)
• 圣卢西亚 (主权实体)
• 圣基茨和尼维斯 (主权实体)
• 英国 (圣卢西亚 (主权实体))
• 乌拉圭 (主权实体)
• 巴拉圭 (主权实体)
• 圣文森特和格林纳丁斯 (主权实体)
• 圣卢西亚 (主权实体)
試合数
16試合
ゴール数
30点
(得失点差 1.8845)
得点王
セント・ルシア (3点)
セント・ルシア (3点)
< 2008 2012 >
カリビアンカップ2010
大会概要
開催国
マルティニーク (5回目)
日程
2010年10月2日 - 12月5日
チーム数
8 10支
出場地数
8 10地
大会結果
優勝
ジャマイカ (5回目)
準優勝
グアドループ
3位
キューバ
4位
セント・ルシア
大観戦数
試合数
16試合
ゴール数
30点
(得失点差 1.8845)
得点王
セント・ルシア (3点)
セント・ルシア (3点)
< 2008 2012 >

Tables

Team	Pld	W	D	L	GF	GA	Pts
キューバ	3	2	1	0	3	0	+3
フレンチ	3	1	2	0	2	1	+1
トリニダード・トバゴ	3	1	0	2	1	1	-2
マルティニーク	3	0	1	2	1	3	-2

2010年11月26日 18:00

トリニダード・トバゴ 0-2 Report

スタッフ・ピエール・アキエ、フォル・ド・フランス

観客数 5,000

主審: スタンリー・ランカスター (ガイアナ)

Images



Visually-Rich Docの多言語分類データセット [Fujinuma+ 2023]

新しいデータでカバーできていない
こと:

- 文書レイアウトの多様性
 - レイアウトが学習データの分布と大きく異なると精度が低下 [Chen+ 2023]

三 ウィキペディア
フリー百科事典

三 カリビアンカップ2010

ページ ノート その他 出典: フリー百科事典「ウィキペディア (Wikipedia)」

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニクで開催された。

予選 [編集]
詳細は「カリビアンカップ2010 予選」を参照

出場国 [編集]
• マルティニク (開催国)
• ジャマイカ (前回優勝)
• キューバ (予選突破)
• グレナダ (予選突破)
• トリニダード・トバゴ (予選突破)
• グアドループ (予選突破)
• アンティグア・バーブーダ (予選突破)
• ガイアナ (予選突破)

グループステージ [編集]
時間は全て (UTC-4).

グループH [編集]

Team	Pld	W	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニク	3	0	1	2	1	3	-2	1

2010年11月26日
18:00
トリニダード・トバゴ  0-2 Report

スタッド・ビエール＝アキケー、フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

E.g., InfoBox
always at upper
right side

カリビアンカップ2010

大会概要

開催国 マルティニク

日程 2010年10月2日 - 12月5日

チーム数 8 (CFU連盟)

開催地数 8 (8都市)

大会結果

優勝 ジャマイカ (5回目)

準優勝 グアドループ

3位 キューバ

4位 グレナダ

大会統計

試合数 16試合

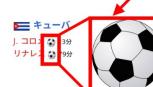
得点数 30点 (試合平均 1.88点)

得点王 キットソン・ペイン
デーン・リチャード (3点)

< 2008 2012 >

Tables

Images

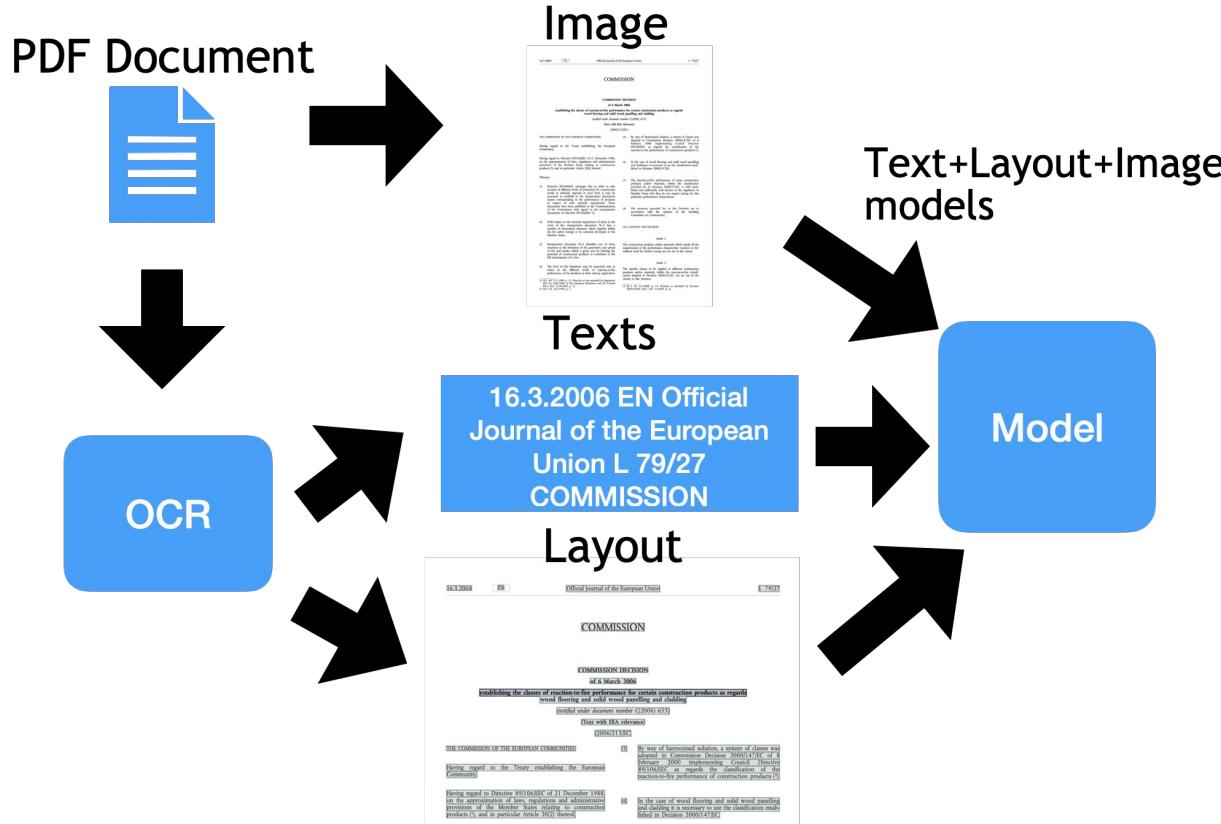


24

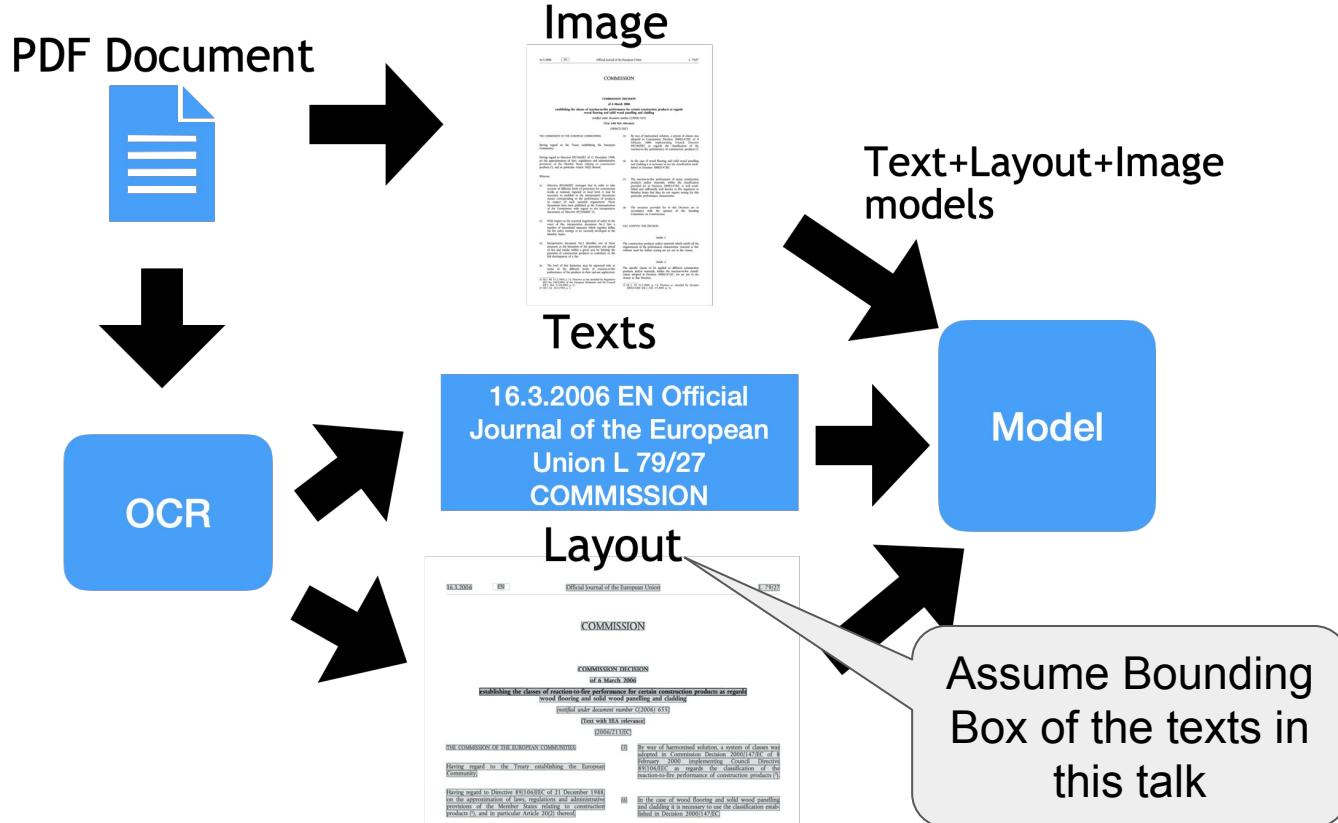
背景: Layout-Aware Modelsについて

- Visually-Rich Docsの処理プロセス
- Visually-Rich Docsに特化したTransformerモデル

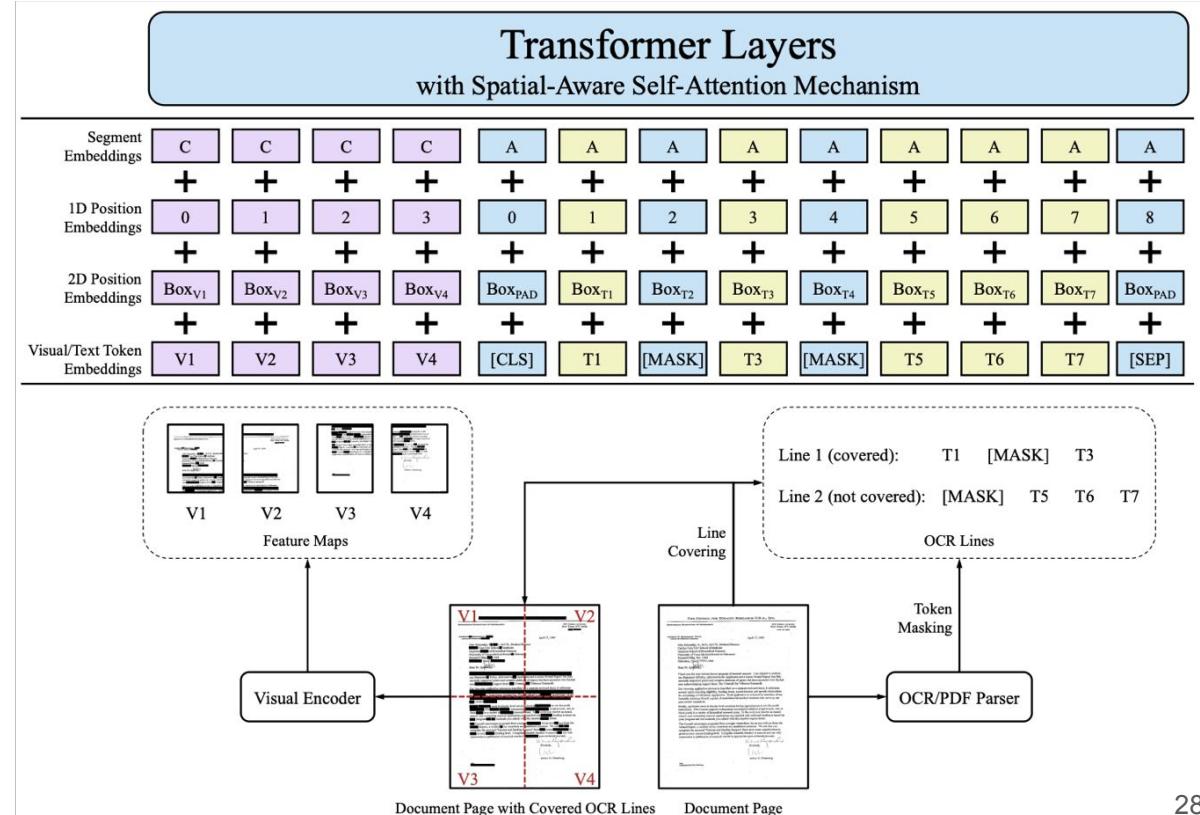
Visually-Rich Documentの処理プロセス



Visually-Rich Documentの処理プロセス

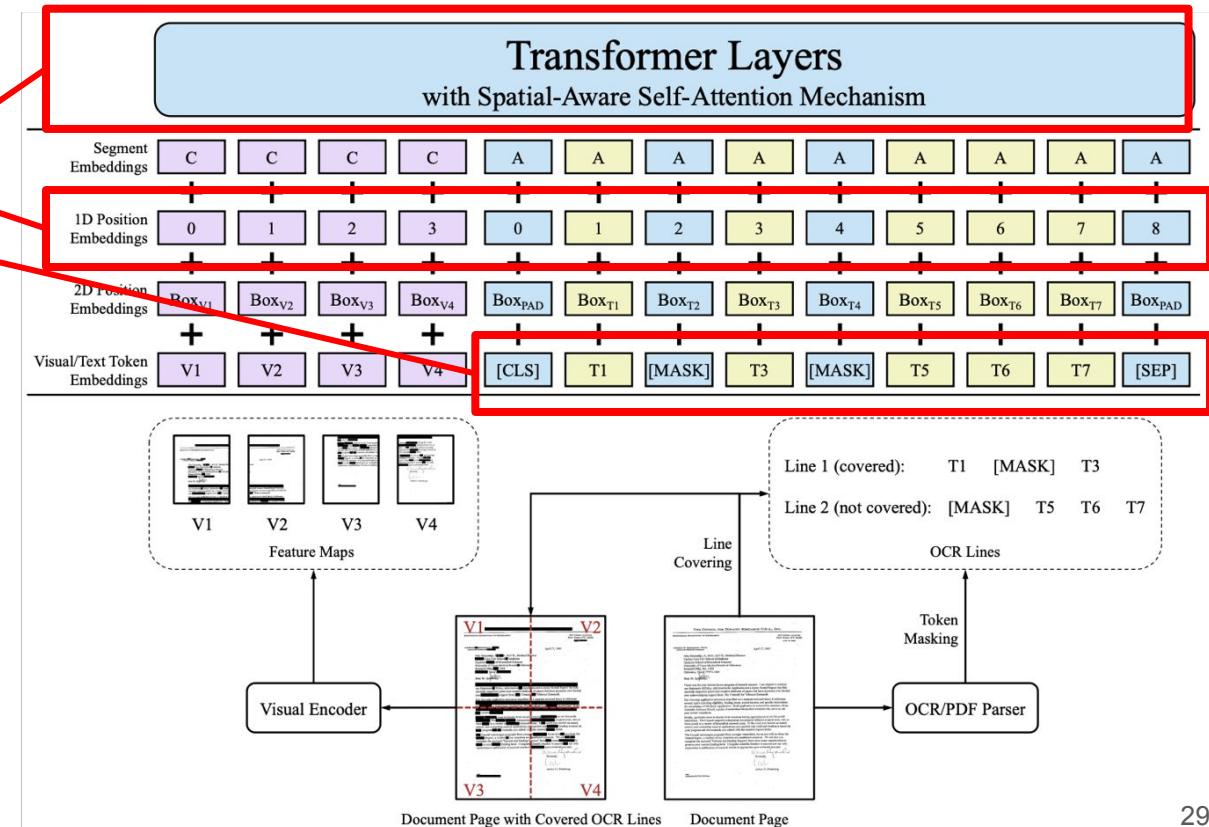


数少ない多言語Visually-Rich Doc Model: LayoutXLM [Xu+ 2021]



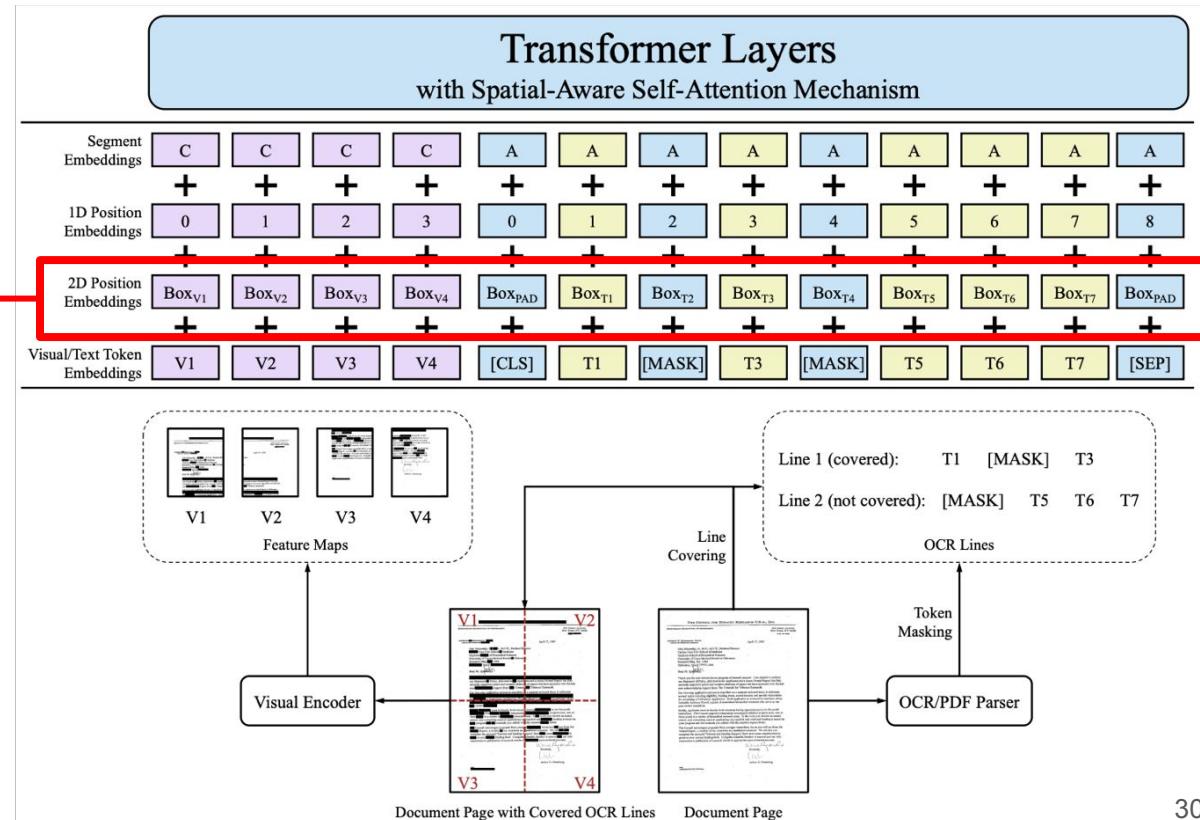
数少ない多言語Visually-Rich Doc Model: LayoutXLM [Xu+ 2021]

ここは通常のTransformer
と同じく
トークン+位置埋め込み

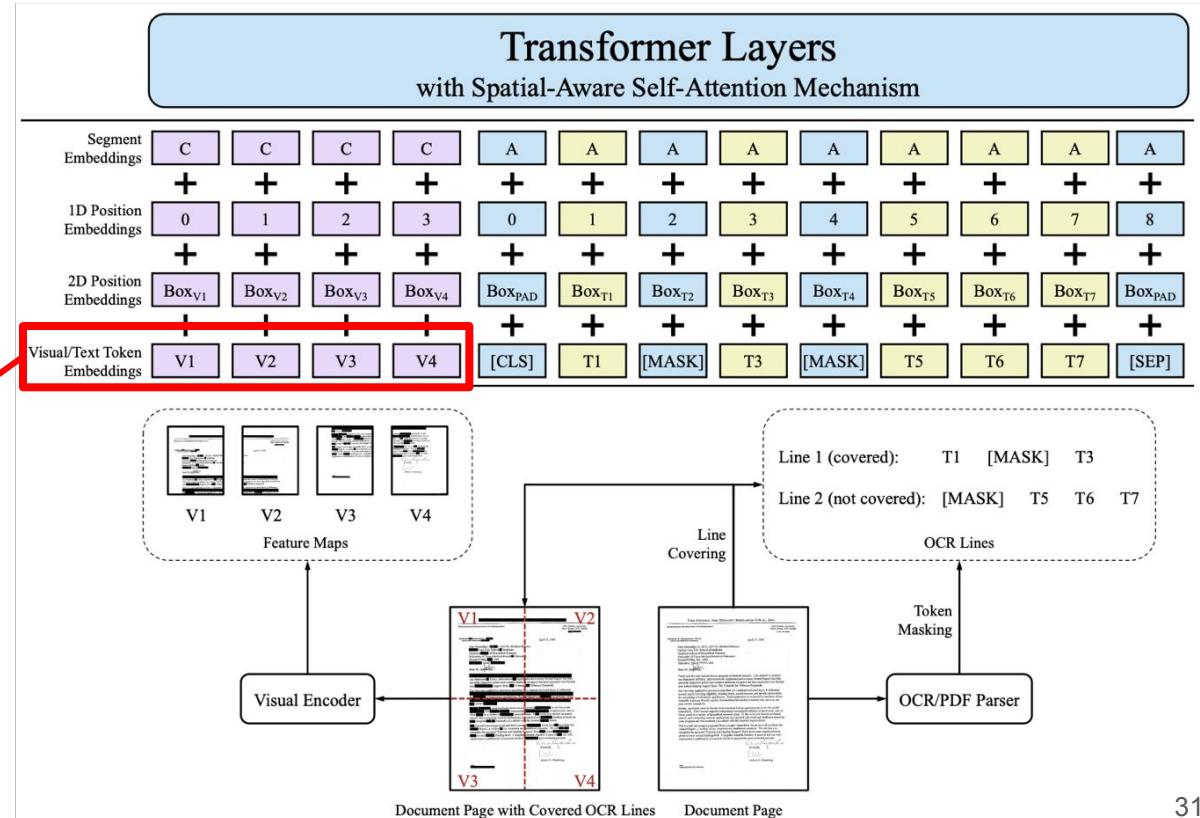


数少ない多言語Visually-Rich Doc Model: LayoutXLM [Xu+ 2021]

ここは通常と異なり
2次元位置埋め込み(文
書レイアウト)



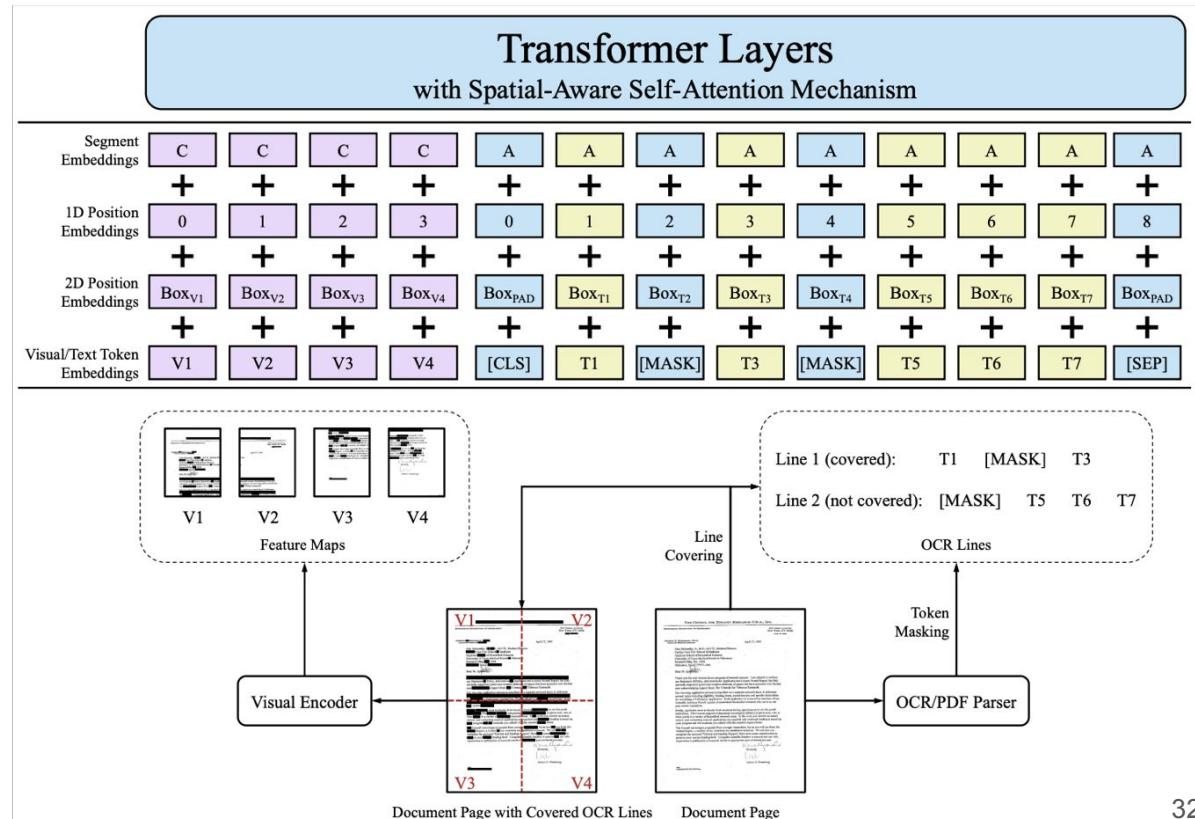
数少ない多言語Visually-Rich Doc Model: LayoutXLM [Xu+ 2021]



ここは通常と異なり
画像の埋め込み



数少ない多言語 Visually-Rich Doc Model: LayoutXLM [Xu+ 2021]



事前学習データは
Common Crawlから抽
出した63言語のPDF

LayoutXLMは多言語には強い

- 学習と推論は同じ言語
- LayoutXLM(マルチモーダル)は精度が言語を問わず安定
 - 多言語PDFの事前学習と入力モダリティが豊富

Models	en	da	de	nl	sv	ro	es	fr	it
InfoXLM	64.98 _{1.7}	63.16 _{1.2}	63.89 _{0.9}	62.82 _{3.6}	64.08 _{1.1}	28.31 _{24.7}	63.2 _{1.7}	65.12 _{0.5}	64.74 _{1.4}
LiLT	61.56 _{2.6}	42.57 _{28.9}	61.48 _{1.3}	59.14 _{2.9}	63.78 _{0.5}	1.01 _{0.3}	63.1 _{1.7}	42.04 _{30.6}	62.84 _{0.7}
LayoutXLM	65.67 _{0.5}	65.17 _{0.7}	65.09 _{0.5}	65.07 _{0.2}	64.76 _{1.0}	64.15 _{1.1}	65.25 _{0.3}	65.36 _{0.7}	65.22 _{0.3}
Donut	29.29	31.69	26.15	25.66	21.94	19.28	24.01	30.92	33.97
Models	pt	pl	bg	cs	hu	fi	el	et	Avg
InfoXLM	64.01 _{1.5}	61.12 _{0.8}	14.23 _{0.1}	40.99 _{34.9}	58.84 _{0.9}	63.46 _{1.0}	63.95 _{0.7}	60.37 _{0.8}	56.89
LiLT	58.10 _{2.4}	58.85 _{0.5}	1.55 _{2.1}	37.60 _{31.8}	39.27 _{33.8}	61.75 _{0.6}	60.45 _{1.8}	59.26 _{0.9}	49.08
LayoutXLM	64.26 _{0.2}	63.26 _{0.7}	63.67 _{0.6}	63.6 _{0.3}	63.87 _{0.8}	63.52 _{1.1}	62.19 _{0.3}	63.43 _{0.2}	64.32
Donut	26.87	22.27	20.03	19.30	23.77	25.70	32.83	22.04	25.60

LayoutXLMは言語間転移が難しい

- 学習と推論は異なる言語
- InfoXLM(テキストのみ)とLayoutXLMの精度に差あり
 - 特に学習言語である英語から離れているウラル系言語で(フィンランド語等)

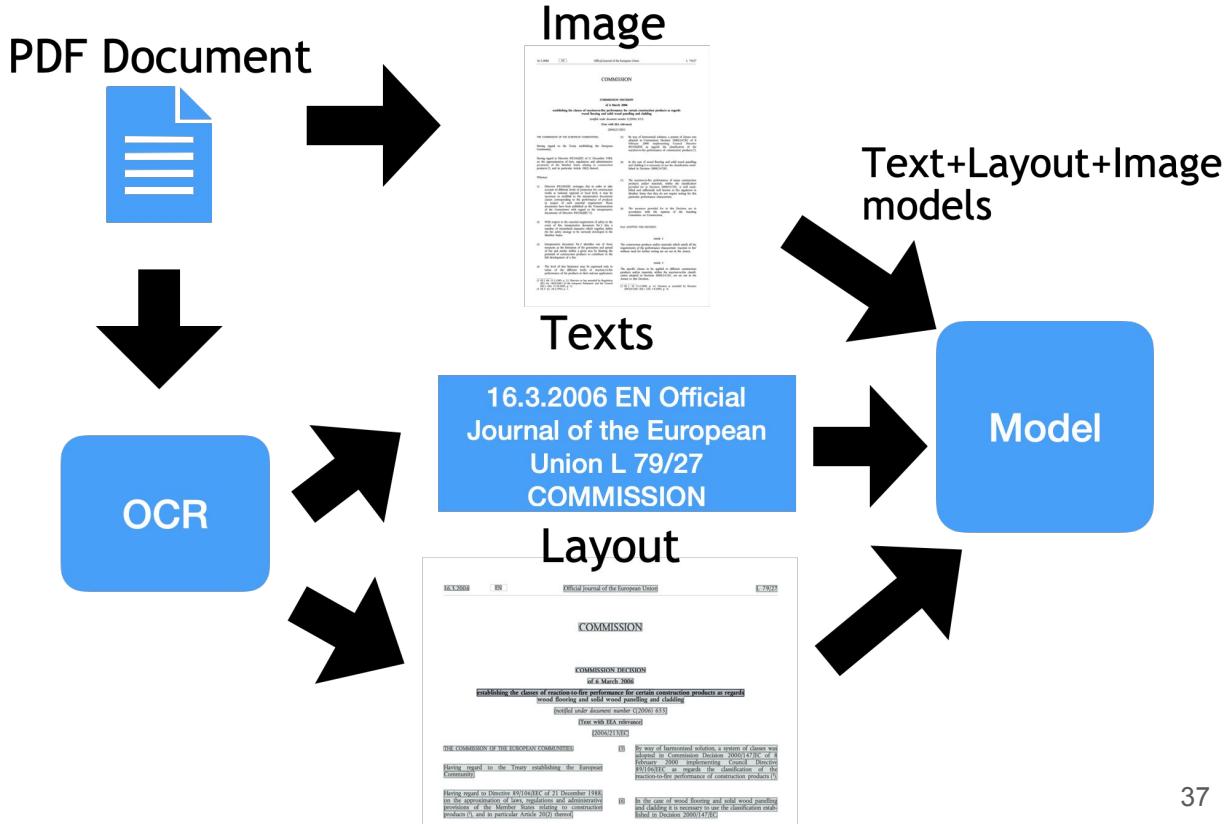
Models	da	de	nl	sv	ro	es	fr	it	pt
InfoXLM	51.28 _{1.3}	53.27 _{1.3}	46.47 _{1.4}	47.91 _{2.5}	48.73 _{2.8}	52.63 _{2.4}	52.25 _{2.6}	47.75 _{2.3}	47.98 _{1.0}
LiLT	43.94 _{6.1}	44.30 _{5.8}	38.75 _{5.4}	42.11 _{7.3}	43.32 _{5.7}	47.41 _{4.6}	43.96 _{6.3}	45.43 _{3.4}	42.99 _{5.5}
LayoutXLM	51.29 _{1.7}	46.26 _{1.5}	46.49 _{2.9}	47.75 _{1.5}	50.15 _{2.1}	52.35 _{1.1}	52.50 _{0.7}	49.33 _{1.6}	48.46 _{1.7}
Donut	14.03	12.19	13.49	14.86	11.25	12.17	10.63	11.00	13.01
Models	pl	bg	cs	hu	fi	el	et	Avg	
InfoXLM	41.62 _{0.6}	45.78 _{2.3}	46.35 _{2.2}	45.74 _{3.4}	42.86 _{3.4}	34.87 _{2.4}	41.78 _{3.7}	46.70	
LiLT	35.49 _{5.3}	40.77 _{6.9}	37.28 _{8.0}	39.03 _{6.8}	34.02 _{7.0}	27.17 _{4.1}	34.41 _{7.1}	40.02	
LayoutXLM	41.28 _{2.7}	47.31 _{1.3}	42.32 _{2.2}	39.36 _{0.9}	31.85 _{1.5}	27.15 _{1.4}	38.37 _{1.8}	44.51	
Donut	12.33	9.32	13.66	8.97	12.25	9.73	16.19	12.19	

今後の課題

- 画像を含むマルチモーダルモデルにおいてより良い言語間転移
- 文書分類問題以外での多言語データセットの構築

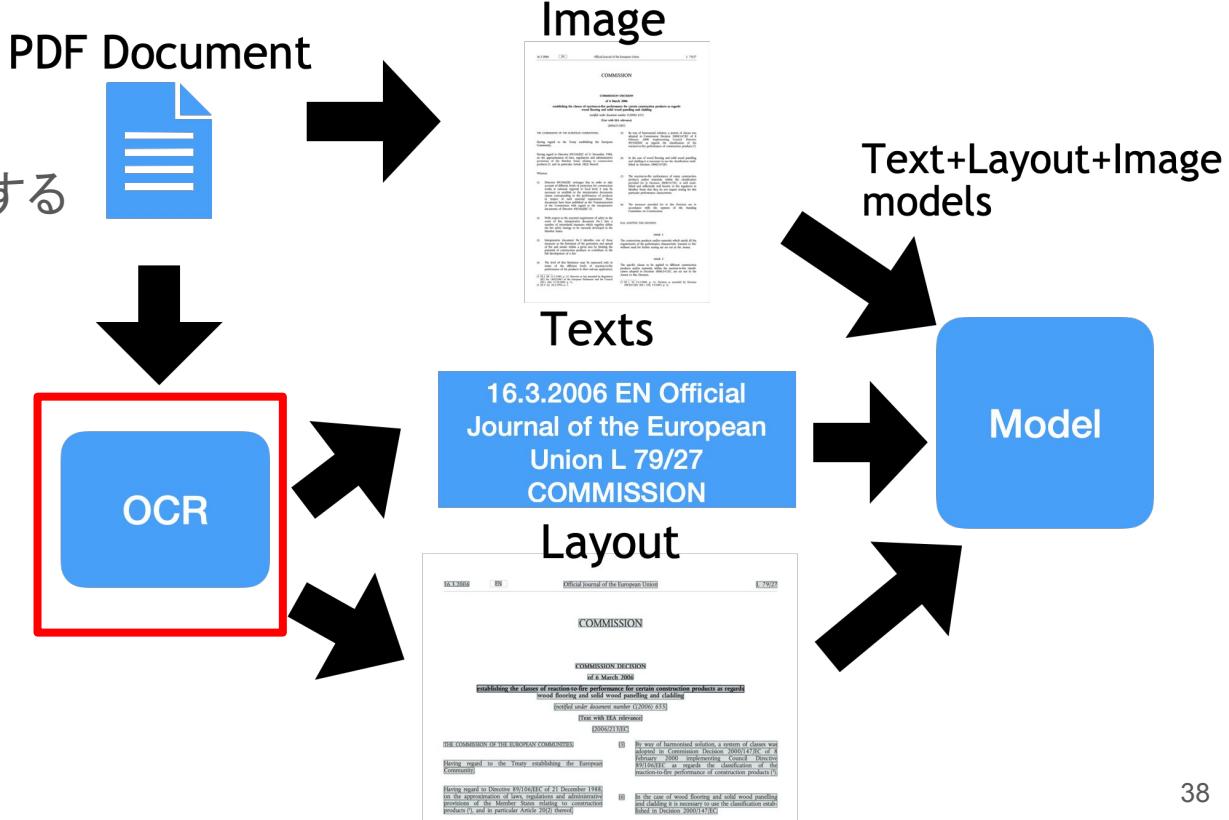
Visually-Rich Documentを軸として見た Vision-Languageモデル(VLM)

そもそもなぜVLMがVisually-Rich Docに対して重要なか？



そもそもなぜVLMがVisually-Rich Docに対して重要なか？

- OCR非依存
- OCRは言語依存
 - 多言語にスケールする上でボトルネック



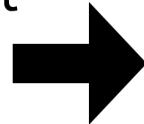
そもそもなぜVLMがVisually-Rich Docに対して重要なか？

- OCR非依存
- OCRは言語依存
 - 多言語にスケールする上でボトルネック
 - 推論時にも遅延

Time (sec/img)



PDF Document



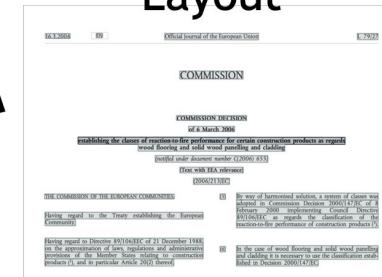
Image



Texts

16.3.2006 EN Official Journal of the European Union L 79/27 COMMISSION

Layout



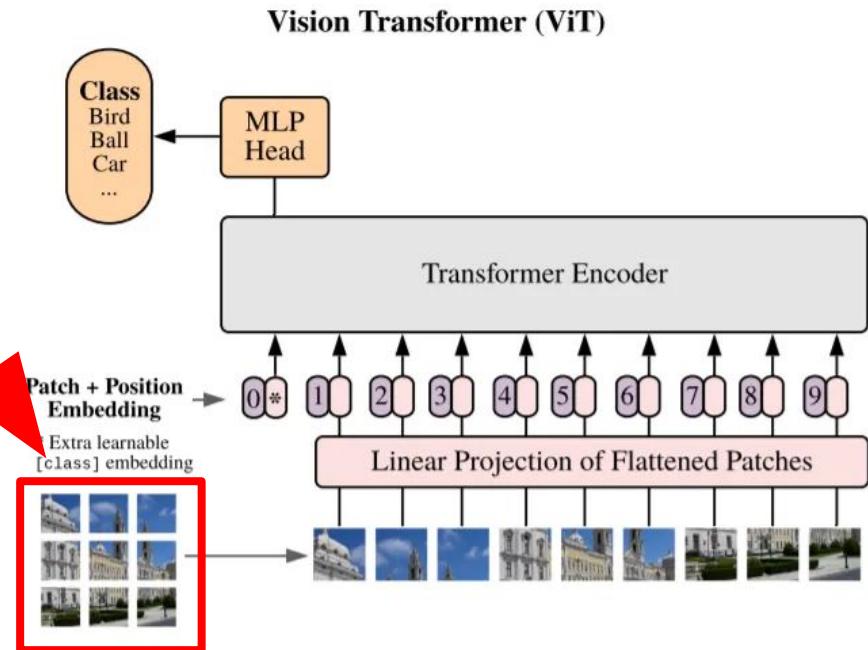
Text+Layout+Image models

Model

[Kim+ 2021]

最近までのVisual-Rich Docから見たVLMの問題点

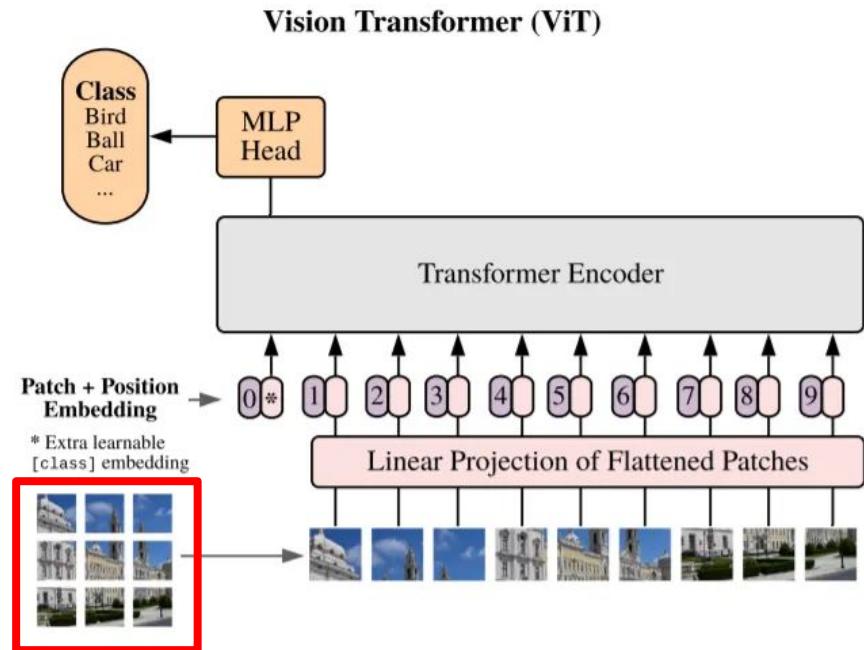
- 入力画像の解像度が低い
 - 固定長の**画像パッチ**を入力したい
 - 画像サイズ(解像度)Resizeする
 - ViTの224x224が(少し前まで?)主流
 - LayoutXLMも224x224
-



[Dosovitskiy+ 2021]より引用

最近までのVisual-Rich Docから見たVLMの問題点

- 入力画像の解像度が低い
 - 固定長の画像パッチを入力したい
 - 画像サイズ(解像度)Resizeする
 - ViTの224x224が(少し前まで?)主流
 - LayoutXLMも224x224
- 画像はデータ量が多い・密
 - 建物の画像は問題ない
 - 文書の画像になると文字が潰れる



[Dosovitskiy+ 2021]より引用

解像度の異なる文書画像

This screenshot shows a low-resolution document from Wikipedia about the Caribbean Cup 2010. It includes:

- A header section titled "カリビアンカップ 2010".
- A detailed schedule table for the tournament.
- A "Group Stage" table showing results for each team: Jamaica, Trinidad & Tobago, Grenada, and Barbados.
- A "Group H" table showing results for the final group stage.
- Footnotes at the bottom providing context about the tournament's name and history.

224x224

This screenshot shows a high-resolution version of the same document from Wikipedia. The differences are:

- The text is much clearer and easier to read.
- The tables are sharper, showing more detail in the cells.
- The overall image quality is significantly better, making it easier to discern small details like the names of the countries and their abbreviations.

448x448

解像度の異なる文書画像



224x224

...
 Wikipedia
 フリー百科事典
 ...

カリビアンカップ2010

ページ ノート その他 出典: フリー百科事典「ウィキペディア (Wikimedia)」

カリビアンカップ2010は、カリビアンカップの2010年に開催された大会である。2011 CONCACAFゴールドカップの予選も兼ねており、2010年10月2日から12月5日までマルティニークで開催された。

予選 [編集]

詳細は「カリビアンカップ2010 予選」を参照

出場国 [編集]

- マルティニーク (開催国)
- ジャマイカ (開催国)
- キューバ (予選突破)
- グレナダ (予選突破)
- トリニダード・トバゴ (予選突破)
- グアドループ (予選突破)
- アンティグア・バーブーダ (予選突破)
- ガイアナ (予選突破)

グループステージ [編集]

時間は全て (UTC-4).

グループH [編集]

Team	Pts	W	D	L	GF	GA	GD	Pts
キューバ	3	2	1	0	3	0	+3	7
グレナダ	3	1	2	0	2	1	+1	5
トリニダード・トバゴ	3	1	0	2	1	3	-2	3
マルティニーク	3	0	1	2	1	3	-2	1

2010年11月26日
18:00
トリニダード・トバゴ 0-2
Reported
キューバ
リオメス 23W
リテレス 90+

スタッド・ピエール＝アギー、フォール・ド・フランス
観客数: 5,000
主審: ス坦リー・ランカスター (ガイアナ)

448x448

解像度の異なる文書画像

This screenshot shows a low-resolution document from the Caribbean Cup 2010. It includes:

- A header section with the tournament name.
- A detailed schedule table for the group stage.
- A results table for the group stage.
- A footer section with tournament details.

This screenshot shows a high-resolution version of the same document. A red box highlights the text "漢字が潰れてしまっている" (The Chinese characters are crushed and distorted) and a red arrow points to the same text in the high-resolution version. The high-resolution version is much clearer and legible.

ページ
カリビ
予選 [結果]
出場国 [結果]
グループステージ [結果]
グループH [結果]
2010年11月26日
18:00
トライニダード・トバゴ vs キューバ
Reported
スタッド・ピエール＝アキター、フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

224x224

448x448

解像度の異なる文書画像



文書画像はタスクや言語によって
必要な解像度が異なる

漢字が潰れてし
まっている

Team	GF	GA	GD	PTS
マルティニーク	3	0	+3	7
ジャマイカ	2	1	+1	5
グレナダ	1	3	-2	3

2010年11月26日
18:00
トロニダード・トバゴ 0-2 キューバ
Reported by: キューバ
スタッド・ビエール＝アキター、フォール・ド・フランス
観客数: 5,000
主審: スタンリー・ランカスター (ガイアナ)

224x224

448x448

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

Table 6: Results on Text-oriented VQA.

Model type	Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
Generalist Models	BLIP-2 (Vicuna-13B)	42.4	-	-	-	-
	InstructBLIP (Vicuna-13B)	50.7	-	-	-	-
	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-	-
	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1	71.3
	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3	75.7
	Qwen-VL-Chat	61.5	62.6	66.3	57.7	70.5
Specialist SOTAs	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2	75.0

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

入力画像サイズ
224x224

Table 6: Results on Text-oriented VQA.

	Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
Generalist Models	BLIP-2 (Vicuna-13B)	42.4	-	-	-	-
	InstructBLIP (Vicuna-13B)	50.7	-	-	-	-
	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-	-
	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1	71.3
	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3	75.7
	Qwen-VL-Chat	61.5	62.6	66.3	57.7	70.5
Specialist SOTAs	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2	75.0

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

Table 6: Results on Text-oriented VQA.

	Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
Generalist Models	448x448					
	BLIP-2 (Vicuna-13B)	42.4	-	-	-	-
	InstructBLIP (Vicuna-13B)	50.7	-	-	-	-
	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-	-
	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1	71.3
	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3	75.7
Specialist SOTAs	Qwen-VL-Chat	61.5	62.6	66.3	57.7	70.5
	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2	75.0

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

Table 6: Results on Text-oriented VQA.

Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
Generalist Models	BLIP-2 (Vicuna-13B)	42.4	-	-	-
	InstructBLIP (Vicuna-13B)	50.7	-	-	-
	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-
	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1
	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3
	Qwen-VL-Chat	61.5	62.6	66.3	57.7
Specialist SOTAs	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2
					75.0

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

Table 6: Results on Text-oriented VQA.

Model type	Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
Generalist models	BLIP-2 (Vicuna-13B)	42.4	-	-	-	-
	StructBLIP (Vicuna-13B)	50.7	-	-	-	-
	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-	-
	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1	71.3
	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3	75.7
	Qwen-VL-Chat	61.5	62.6	66.3	57.7	70.5
Specialist SOTAs	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2	75.0

企業のTech Reportだと解像度が異なるモデルを比較

- AlibabaのQWEN-VL [Bai+ 2023]

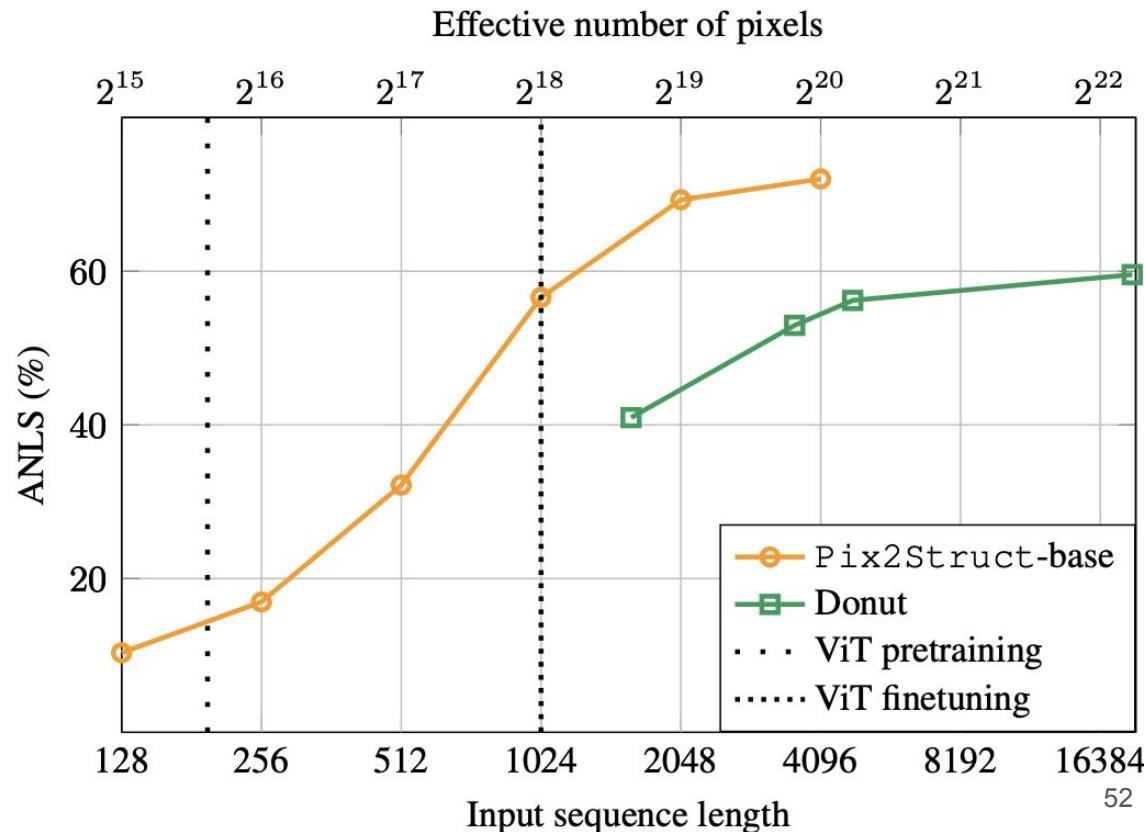
文書画像のQAで解像度が異なるモデルを比較するのは公平かつ有用な知見を得られるのか？

nted VQA.

解像度	Model	TextVQA	DocVQA	ChartQA	AI2D	OCR-VQA
448x448	BLIP-2 (Vicuna-13B)	42.4	-	-	-	-
448x448	InstructBLIP (Vicuna-13B)	50.7	-	-	-	-
最大1024x1024	mPLUG-DocOwl (LLaMA-7B)	52.6	62.2	57.4	-	-
448x448	Pic2Struct-Large (1.3B)	-	76.6	58.6	42.1	71.3
448x448	Qwen-VL (Qwen-7B)	63.8	65.1	65.7	62.3	75.7
448x448	Qwen-VL-Chat	61.5	62.6	66.3	57.7	70.5
Specialist SOTAs	PALI-X-55B (Single-task fine-tuning, without OCR Pipeline)	71.44	80.0	70.0	81.2	75.0

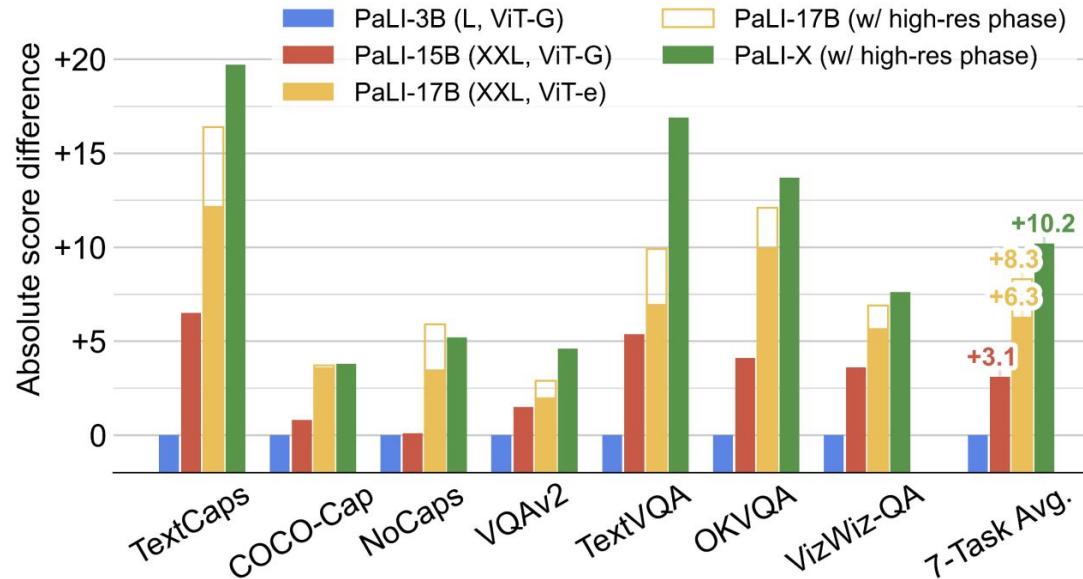
Pix2structの貢献 [Appendix of Lee+ 2022]

- 横軸が入力解像度
- 縦軸が精度
- DocVQAにおいて、入力画像サイズの重要性が一目瞭然



PaLI-Xの貢献 [Chen+ 2023]

- Text Transformer(33B)とVision Transformer(22B)両方スケール
- 事前学習で徐々に画像の解像度を上げていている
 - 224×224→448×448→672×672→756×756



研究部分の総括と今後の課題

- 文書画像はタスクや言語によって必要な解像度が異なる
- 企業のTech Reportは半分宣伝資料？
- マルチモーダルモデルでの言語間転移を改善するには？
- 多言語データセットはまだまだ不足