La Rochelle Université

D'ici, on voit + loin !

EAML: Ensemble Self-Attention-based Mutual Learning Network for Document Image Classification

Souhail Bakkali, Mickaël Coustaty, Zuheng Ming, Marçal Rusiñol, Oriol Ramos Terrades

14/10/2022

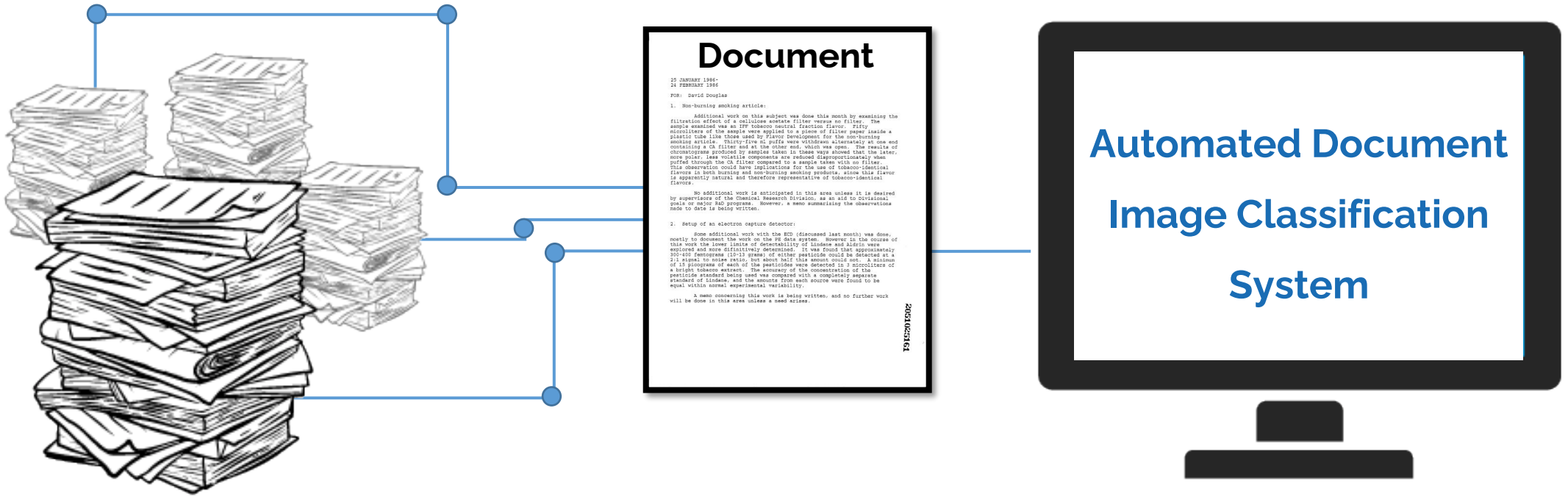
La Rochelle Université

Informatique Image Interaction
Univ. La Rochelle

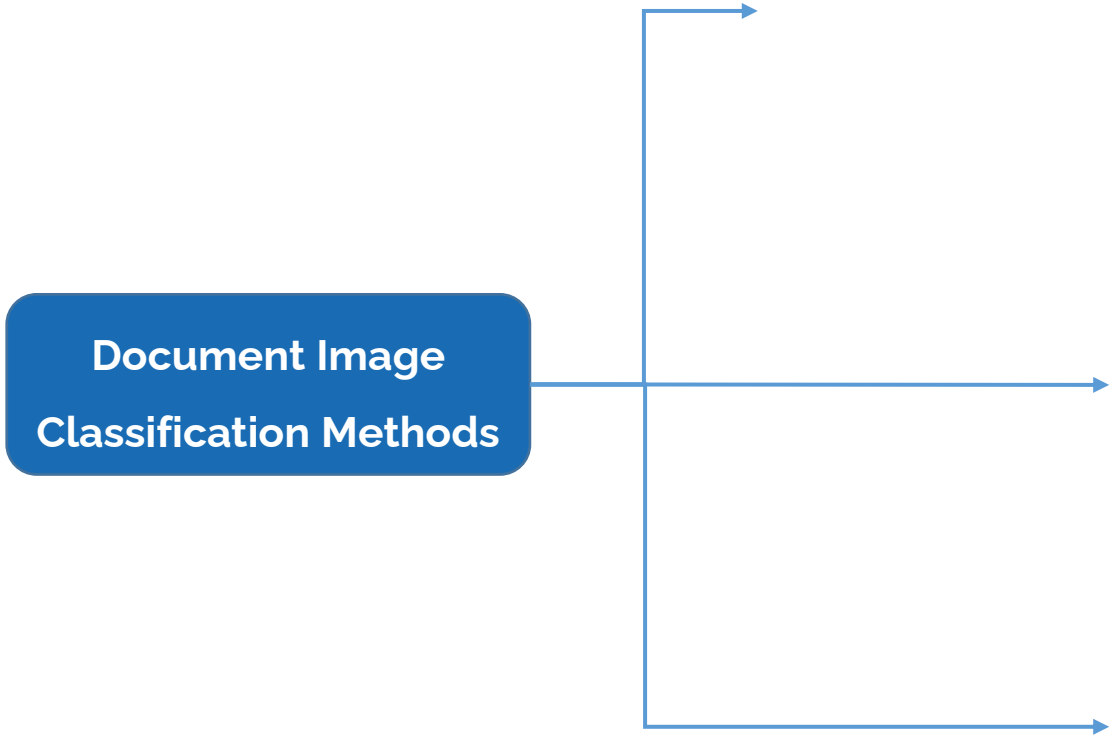
Centre de Visió per Computador

- PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

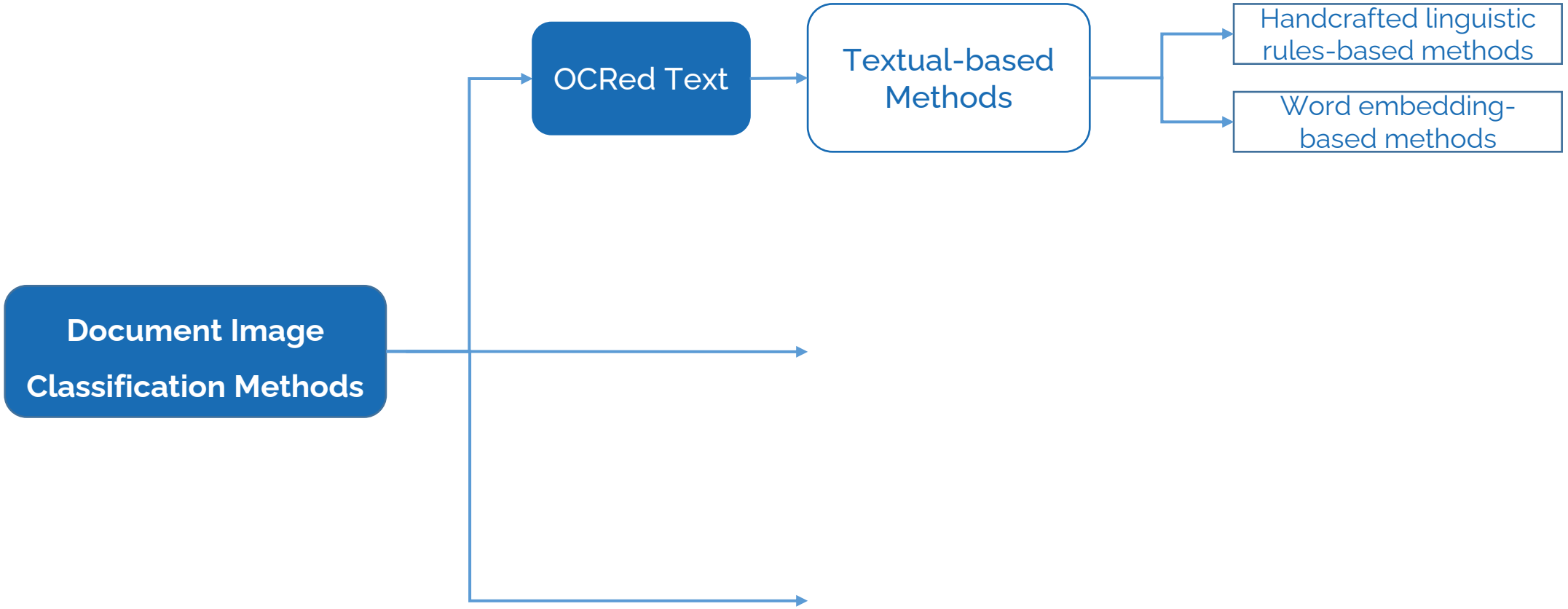
Why do we need Multimodal Document Image Classification ?



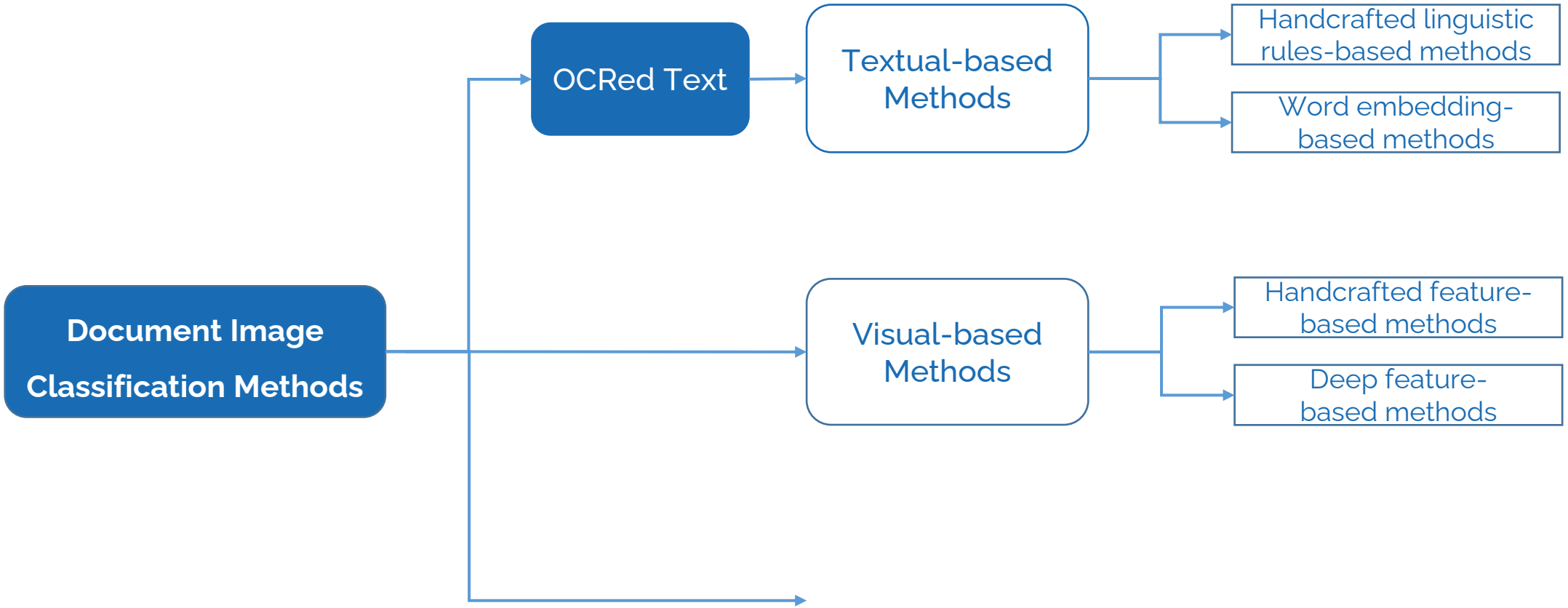
The Existing Document Image classification Methods



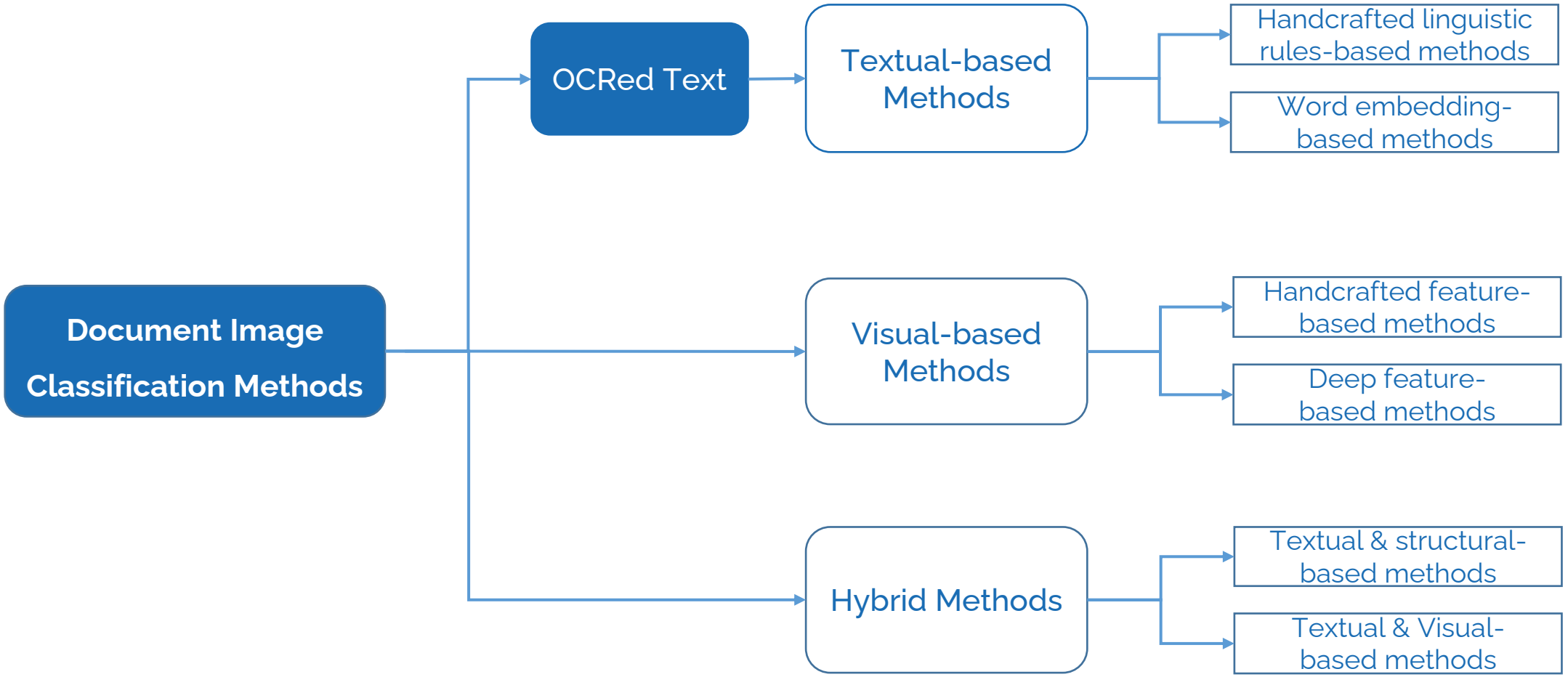
The Existing Document Image classification Methods



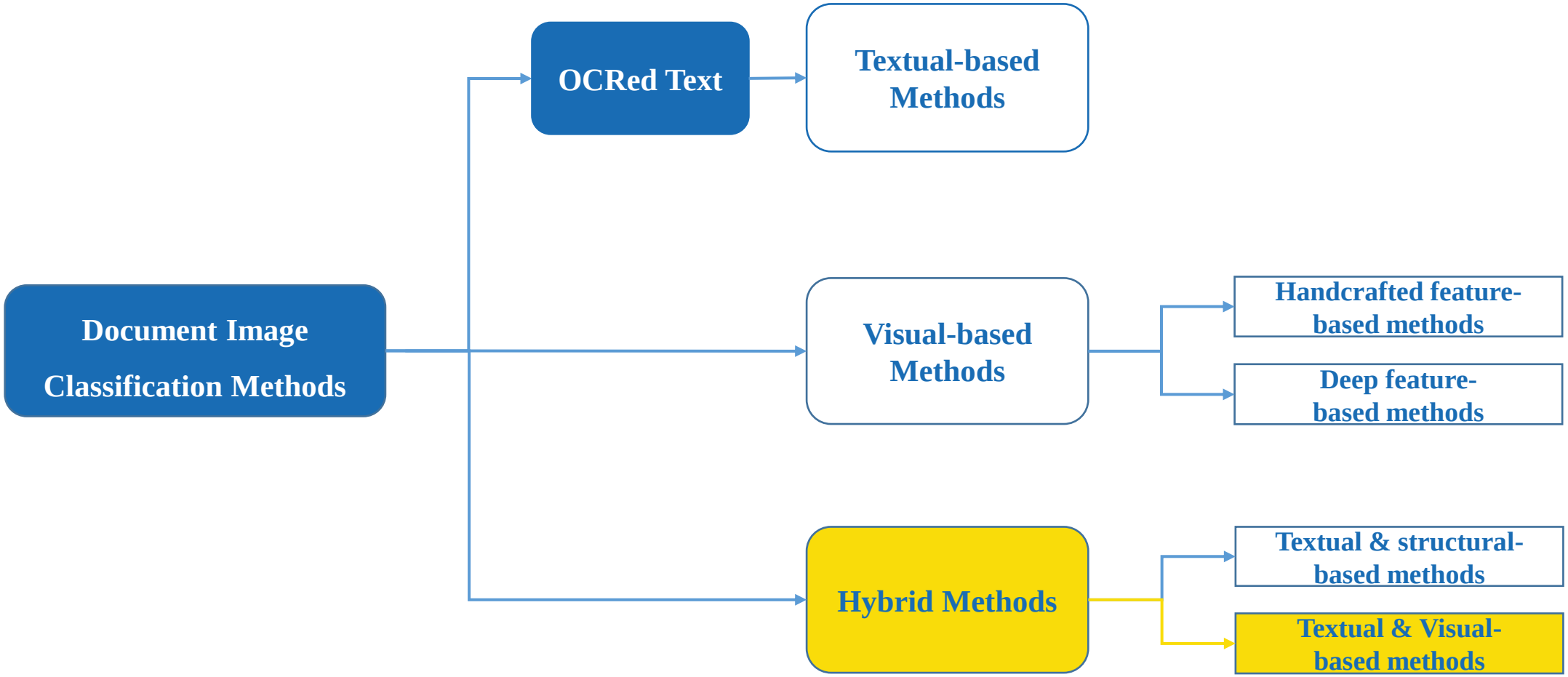
The Existing Document Image classification Methods



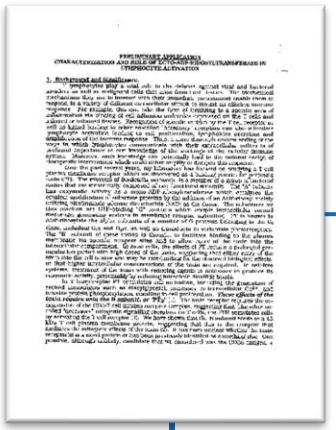
The Existing Document Image classification Methods



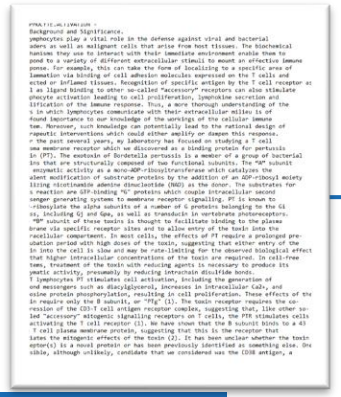
The Existing Document Image classification Methods



The Need for a Supervised Hybrid « Multi-Modal » Method



OCRed Text



Vision Modality

Language Modality

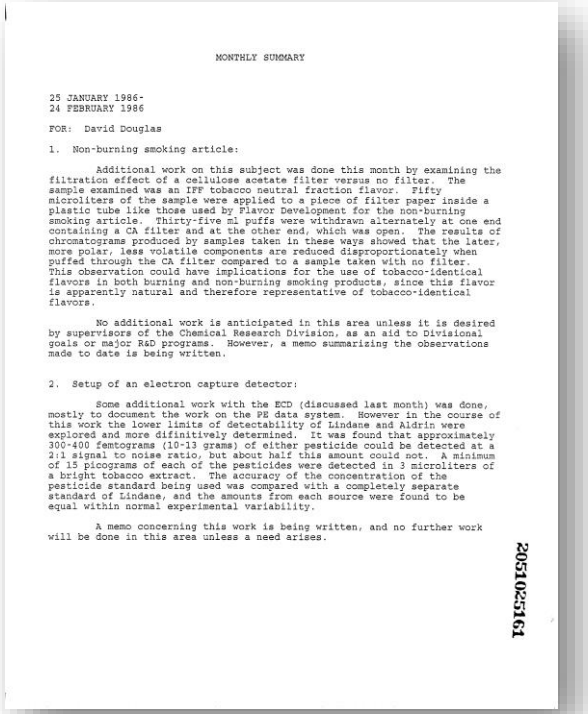
Cross-Modal Learning

Output Embeddings

Predictions

Challenges: High « Intra-class » Structural Variability
Low « Inter-class » Discrimination

Category « A »



MONTHLY SUMMARY

25 JANUARY 1986-
24 FEBRUARY 1986

FOR: David Douglas

1. Non-burning smoking article:

Additional work on this subject was done this month by examining the filtration effect of a cellulose acetate filter versus no filter. The sample examined was an IFF tobacco neutral fraction flavor. Fifty microliters of the sample were applied to a piece of filter paper inside a plastic tube like those used by Flavor Development for the non-burning smoking article. Thirty-five ml puffs were withdrawn alternately at one end containing a CA filter and at the other end, which was open. The results of chromatograms produced by samples taken in these ways showed that the later, more polar, less volatile components are reduced disproportionately when puffed through the CA filter compared to a sample taken with no filter. This observation could have implications for the use of tobacco-identical flavors in both burning and non-burning smoking products, since this flavor is apparently natural and therefore representative of tobacco-identical flavors.

No additional work is anticipated in this area unless it is desired by supervisors of the Chemical Research Division, as an aid to Divisional goals or major R&D programs. However, a memo summarizing the observations made to date is being written.

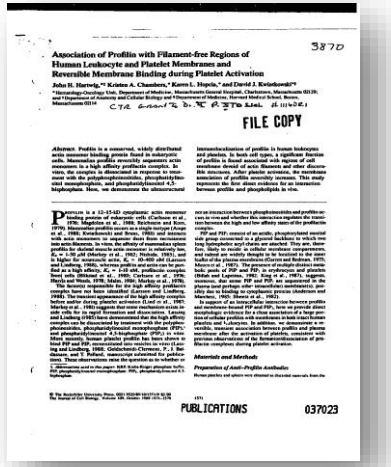
2. Setup of an electron capture detector:

Some additional work with the ECD (discussed last month) was done, mostly to document the work on the PE data system. However in the course of this work the lower limits of detectability of Lindane and Aldrin were explored and more difinitively determined. It was found that approximately 300-400 femtograms (10-13 grams) of either pesticide could be detected at a 2:1 signal to noise ratio, but about half this amount could not. A minimum of 15 picograms of each of the pesticides were detected in 3 microliters of a bright tobacco extract. The accuracy of the concentration of the pesticide standard being used was compared with a completely separate standard of Lindane, and the amounts from each source were found to be equal within normal experimental variability.

A memo concerning this work is being written, and no further work will be done in this area unless a need arises.

2051025161

Category « A »



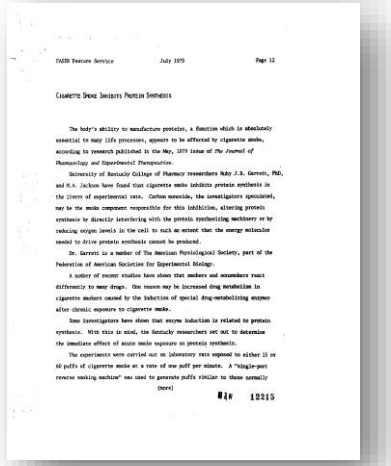
3870

Association of Profilin with Filament-free Regions of Human Leukocyte and Platelet Membranes and Reversible Membrane Binding during Platelet Activation

FILE COPY

PUBLICATIONS 037023

Category « B »



FASEB Feature Service July 1979 Page 12

CIGARETTE SMOKE INHIBITS PROTEIN SYNTHESIS

The body's ability to manufacture proteins, a function which is absolutely essential to many life processes, appears to be affected by cigarette smoke, according to research published in the May, 1979 issue of *The Journal of Pharmacology and Experimental Therapeutics*.

University of Kentucky College of Pharmacy researchers Ruby J.S. Garrett, PhD, and H.A. Jackson have found that cigarette smoke inhibits protein synthesis in the livers of experimental rats. Carbon monoxide, the investigators speculated, may be the smoke component responsible for this inhibition, altering protein synthesis by directly interfering with the protein synthesizing machinery or by reducing oxygen levels in the cell to such an extent that the energy molecules needed to drive protein synthesis cannot be produced.

Dr. Garrett is a member of The American Physiological Society, part of the Federation of American Societies for Experimental Biology.

A number of recent studies have shown that smokers and nonsmokers react differently to many drugs. One reason may be increased drug metabolism in cigarette smokers caused by the induction of special drug-metabolizing enzymes after chronic exposure to cigarette smoke.

Some investigators have shown that enzyme induction is related to protein synthesis. With this in mind, the Kentucky researchers set out to determine the immediate effect of acute smoke exposure on protein synthesis.

The experiments were carried out on laboratory rats exposed to either 15 or 60 puffs of cigarette smoke at a rate of one puff per minute. A "single-port reverse smoking machine" was used to generate puffs similar to those normally [here]

12215

Our Proposed EAML Network

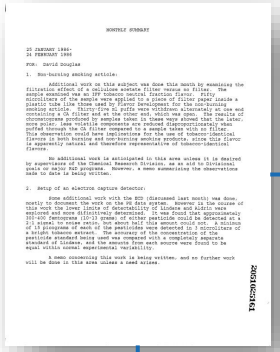
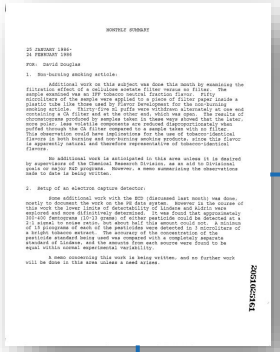


Image Branch
(with attention module)

OCRed Text

Text Branch
(with attention module)

Our Proposed EAML Network



OCRed Text

MONTHLY SUMMARY

25 JANUARY 1986-
24 FEBRUARY 1986

FOR: David Douglas

1. Non-burning smoking article:

Additional work on this subject was done this month by examining the filtration effect of a cellulose acetate filter versus no filter. The sample examined was an IFF tobacco neutral fraction flavor. Fifty microliters of the sample were applied to a piece of filter paper inside a glass tube like those used by Flavor Development for the non-burning smoking article. Thirty-five 81 puffs were withdrawn alternately at one end containing a CA filter and at the other end, which was open. The results of chromatograms produced by samples taken in these ways showed that the later, more polar, less volatile components are reduced disproportionately when puffed through the CA filter compared to a sample taken with no filter. This observation could have implications for the use of tobacco-identical flavors in both burning and non-burning smoking products, since this flavor is apparently natural and therefore representative of tobacco-identical flavors.

No additional work is anticipated in this area unless it is desired by supervisors of the Chemical Research Division, as an aid to Divisional goals, or major R&D programs. However, a memo summarizing the observations made to date is being written.

2. Setup of an electron capture detector:

Some additional work with the ECD (discussed last month) was done, mostly to document the work on the PE data system. However in the course of this work the lower limits of detectability of Lindane and Aldrin were explored and more definitively determined. It was found that approximately 300-400 femtograms (10-13 grams) of either pesticide could be detected at a 2:1 signal to noise ratio, but about half this amount could not. A minimum of 15 picograms of each of the pesticides were detected in 3 microliters of bright tobacco extract. The accuracy of the concentration of the pesticide standard being used was compared with a completely separate standard of Lindane, and the amounts from each source were found to be equal within normal experimental variability.

A memo concerning this work is being written, and no further work will be done in this area unless a need arises.

FITSZOTSO=

MONTHLY SUMMARY

25 JANUARY 1986-
24 FEBRUARY 1986

FOR: David Douglas

1. Non-burning smoking article:

Additional work on this subject was done this month by examining the filtration effect of a cellulose acetate filter versus no filter. The sample examined was an IFF tobacco neutral fraction flavor. Fifty microliters of the sample were applied to a piece of filter paper inside a glass tube like those used by Flavor Development for the non-burning smoking article. Thirty-five 81 puffs were withdrawn alternately at one end and at the other end, which was open. The results of chromatograms produced by samples taken in these ways showed that the later, more polar, less volatile components are reduced disproportionately when puffed through the CA filter compared to a sample taken with no filter. This observation could have implications for the use of tobacco-identical flavors in both burning and non-burning smoking products, since this flavor is apparently natural and therefore representative of tobacco-identical flavors.

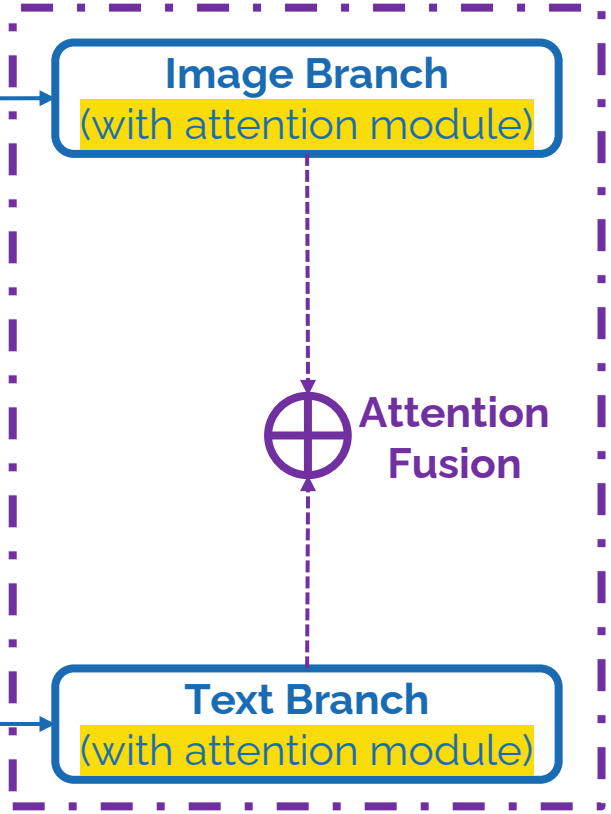
No additional work is anticipated in this area unless it is desired by supervisors of the Chemical Research Division, as an aid to Divisional goals, or major R&D programs. However, a memo summarizing the observations made to date is being written.

2. Setup of an electron capture detector:

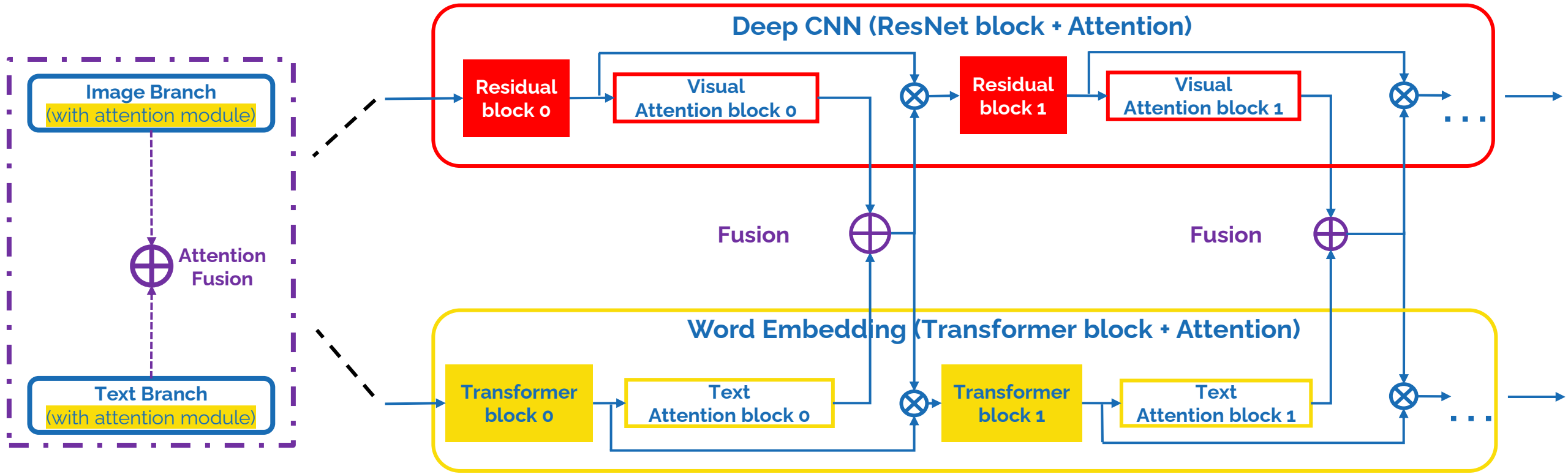
Some additional work with the ECD (discussed last month) was done, mostly to document the work on the PE data system. However in the course of this work the lower limits of detectability of Lindane and Aldrin were explored and more definitively determined. It was found that approximately 300-400 femtograms (10-13 grams) of either pesticide could be detected at a 2:1 signal to noise ratio, but about half this amount could not. A minimum of 15 picograms of each of the pesticides were detected in 3 microliters of bright tobacco extract. The accuracy of the concentration of the pesticide standard being used was compared with a completely separate standard of Lindane, and the amounts from each source were found to be equal within normal experimental variability.

A memo concerning this work is being written, and no further work will be done in this area unless a need arises.

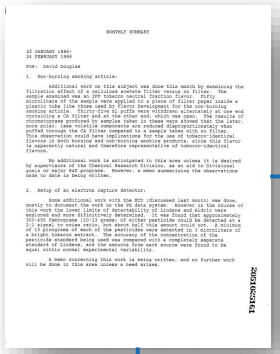
FITSZOTSO=



Self-Attention-based Fusion Module



Our Proposed EAML Network



OCRed Text

Image Branch (with attention module)

Attention Fusion

Text Branch (with attention module)

Predictions (Ps1)

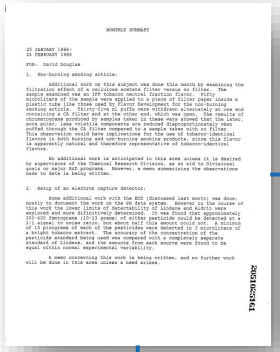
Predictions (Ps2)

Positive Mutual Learning

Image Modality (L1)

Text Modality (L2)

Our Proposed EAML Network



OCRed Text

Image Branch (with attention module)

Text Branch (with attention module)

Attention Fusion

Predictions (Ps1)

Predictions (Ps2)

Positive Mutual Learning

Cross-Modal Fusion

Softmax

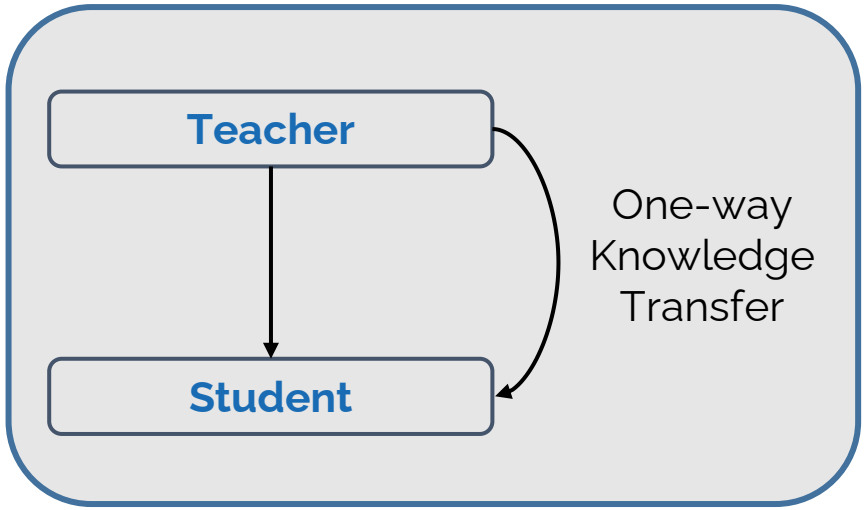
Image Modality (L1)

Fusion Modality (L3)

Text Modality (L2)

Conventional Mutual Learning in Computer Vision

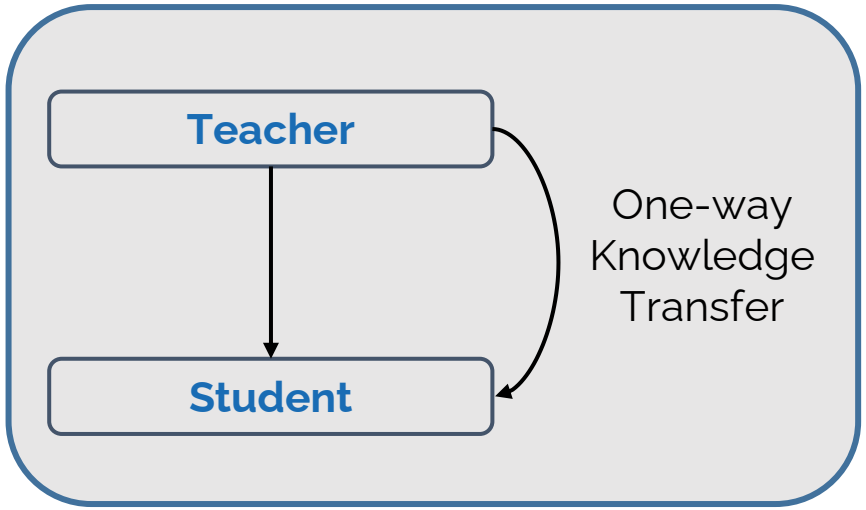
Knowledge Distillation-based approach



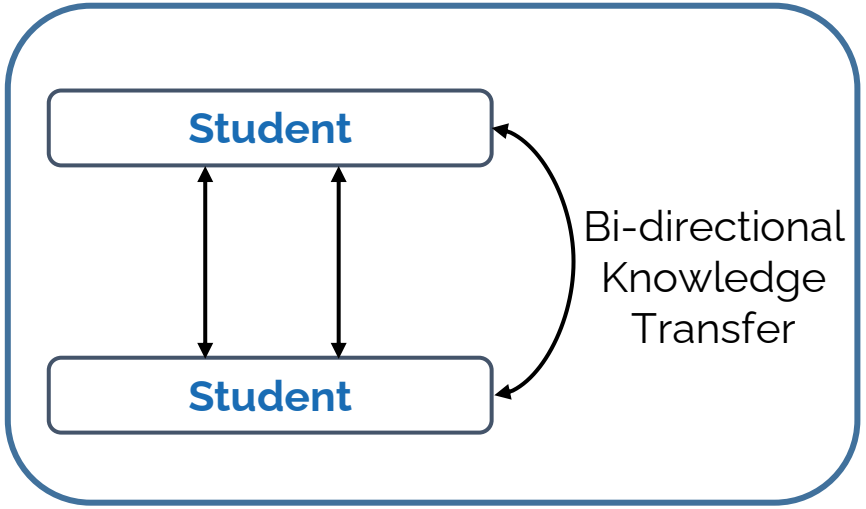
Hinton, Geoffrey E., Oriol Vinyals and J. Dean. "Distilling the Knowledge in a Neural Network." ArXiv abs/1503.02531 (2015): n. pag.

Conventional Mutual Learning in Computer Vision

Knowledge Distillation-based approach



Conventional Mutual Learning-based approach

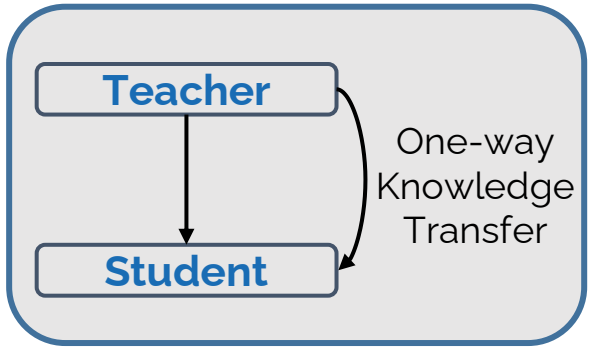


Hinton, Geoffrey E., Oriol Vinyals and J. Dean. "Distilling the Knowledge in a Neural Network." ArXiv abs/1503.02531 (2015): n. pag.

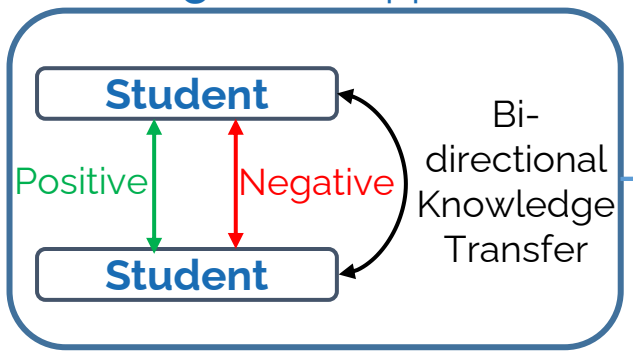
Zhang, Y., T. Xiang, Timothy M. Hospedales And Huchuan Lu. "Deep Mutual Learning." *2018 IEEE/CVF Conference On Computer Vision And Pattern Recognition* (2018): 4320-4328.

Conventional Mutual Learning in Computer Vision: Challenges

Knowledge Distillation-based approach



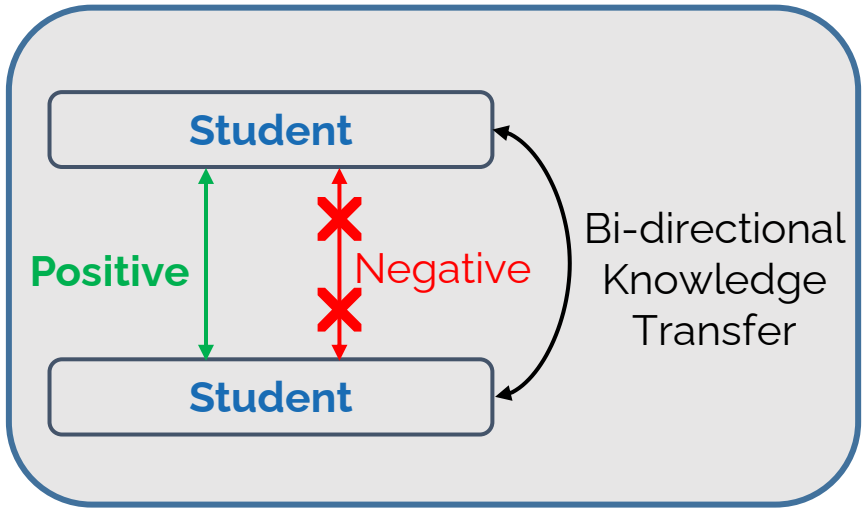
Conventional Mutual Learning-based approach



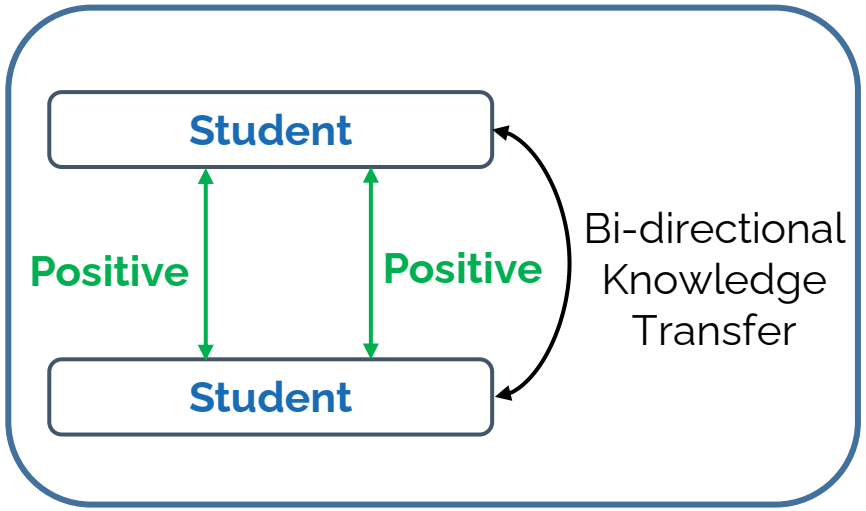
- ✓ Encourages collaborative learning between modalities.
- ✓ Enables to minimize the difference in class probabilities produced by the two modalities
- ✗ Both modalitites can learn both the positive & negative information from one another
- ✗ The introduction of the negative knowledge from one modality to another harms and weakens the ongoing training.

Our Proposed Positive Mutual Learning Strategy

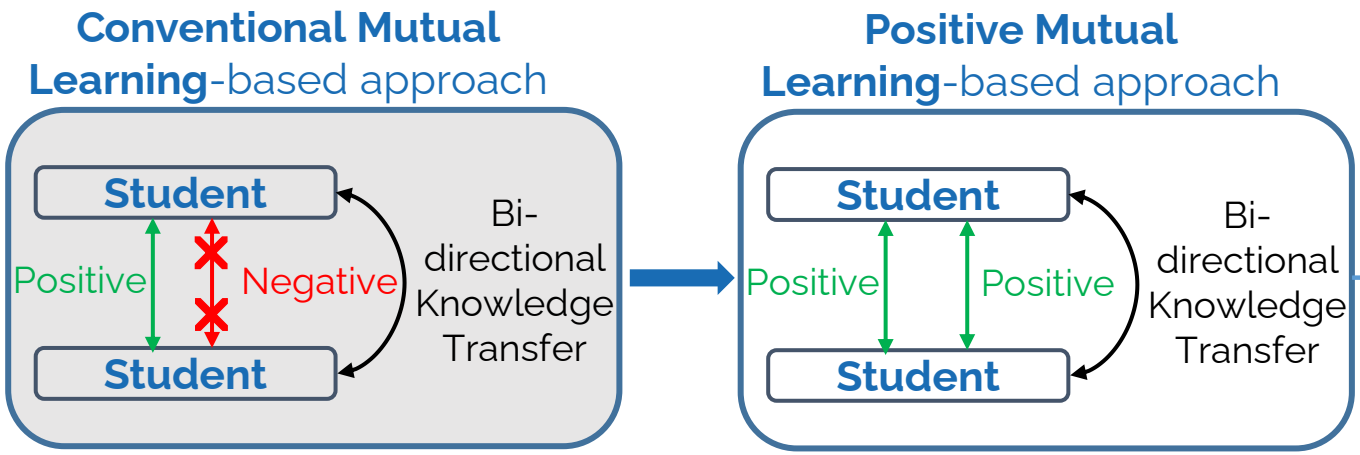
Conventional Mutual Learning-based approach



Positive Mutual Learning-based approach (Ours)



Our Proposed Positive Mutual Learning Strategy



- Enable to learn only the positive knowledge from one modality to another
- Prevent the negative knowledge to be introduced in the ongoing learning of the current modality.
- Improve the quality of the final predictions of the single-modal and cross-modal modalities.

Experimental Setup & Ablation Study

+ Experimental Setup

> Datasets

- RVLCDIP
- Tobacco-3482

+ Ablation Study

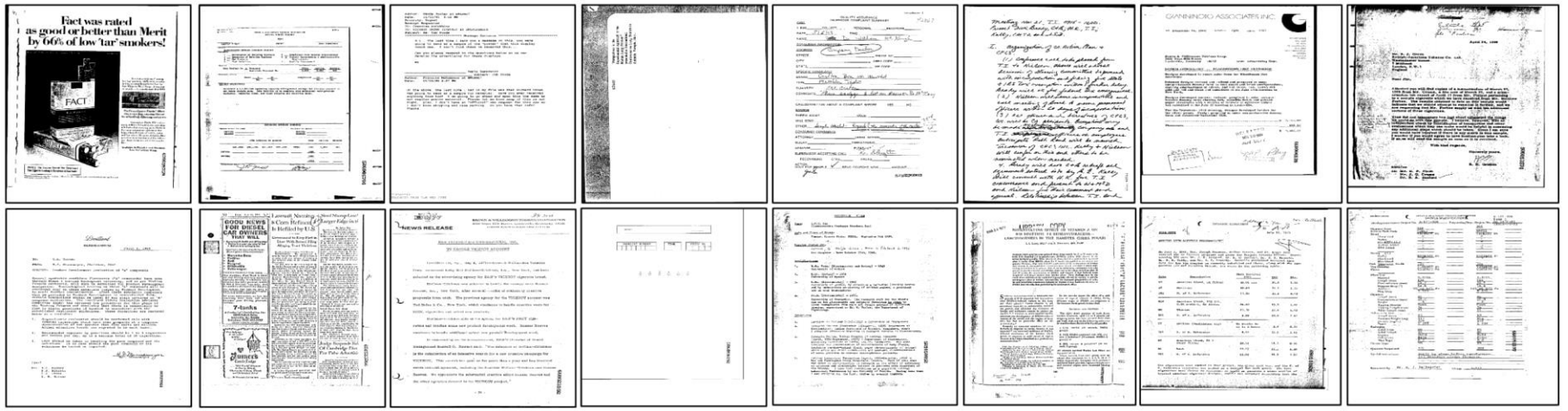
> **Intra**-Dataset Evaluation

> **Inter**-Dataset Evaluation

> **Intra**-Dataset & **Inter**-Dataset Evaluation comparison

Experimental Setup: Datasets

RVLCDIP



- **400,000** gray-scale document images
- Dataset split:
  - Training set: 320,000 images.
  - Validation set: 40,000 images.
  - Test set: 40,000 images.
- Samples of different document classes in the **RVL-CDIP** dataset. From left to right: *Advertisement*, *Budget*, *Email*, *File folder*, *Form*, *Handwritten*, *Invoice*, *Letter*, *Memo*, *News article*, *Presentation*, *Questionnaire*, *Resume*, *Scientific publication*, *Scientific report*, *Specification*.

Experimental Setup: Datasets

Tobacco-3482

- **3482** gray-scale document images.
- Samples of different document classes in the **Tobacco-3482** dataset. The 10 categories are: *ADVE*, *Email*, *Form*, *Letter*, *Memo*, *News*, *Notes*, *Report*, *Resume*, *Scientific*.
- Dataset Split:
  - 80% Training set
  - 10% Validation set
  - 10% Test set.

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

RVLCDIP	Method	Image Modality	Text Modality	Fusion Modality
	Independent Learning	**85.04**	**84.96**	**94.44**

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

RVLCDIP	Method	Image Modality	Text Modality	Fusion Modality
	Independent Learning	**85.04**	**84.96**	**94.44**
	Conventional Mutual Learning	**88.87**	**80.89**	**90.06**

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

RVLCDIP	Method	Image Modality	Text Modality	Fusion Modality
	Independent Learning	**85.04**	**84.96**	**94.44**
	Conventional Mutual Learning	**88.87**	**80.89**	**90.06**
	Positive Mutual Learning (**Ours**)	**90.81**	**88.80**	**96.28**

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

RVLCDIP	Method	Image Modality	Text Modality	Fusion Modality
	Independent Learning	**85.04**	**84.96**	**94.44**
	Conventional Mutual Learning	**88.87**	**80.89**	**90.06**
	Positive Mutual Learning (**Ours**)	**90.81**	**88.80**	**96.28**
	EAML	**97.67**	**97.63**	**97.70**

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

Tobacco-3482	Method	Image Modality	Text Modality	Fusion Modality
	Independent Learning	**96.17**	**96.02**	**96.95**
	Conventional Mutual Learning	**93.69**	**88.82**	**94.84**
	Positive Mutual Learning (**Ours**)	**97.70**	**96.27**	**98.28**
	EAML	**97.99**	**96.27**	**98.57**

Ablation Study : Intra-Dataset Evalutation

Overall Classification Accuracy (%) on the proposed approaches

RVLCDIP

Method	Model	Accuracy(%)
Image	Nicolas *et al.* [7]	89.1
Text		74.6
Multi-modal		90.6
Image	Dauphinee *et al.* [17]	90.24
Text		82.23
Multi-Modal		93.07
Image	Cross-Modal [11]	91.45
Text		84.96
Multi-modal		97.05
Image	$EAML_{Tr-KLD_{Reg}}$ (Ours)	**97.67**
Text		**97.63**
Multi-Modal		**97.70**
Baselines	Harley *et al.* [25]	89.80
	Csurka *et al.* [15]	90.70
	Tensmeyer *et al.* [56]	90.94
	Azfal *et al.* [2]	90.97
	Das *et al.* [16]	91.11
	Das *et al.* [16]	92.21
	Ferrando *et al.* [21]	92.31
	Xu *et al.* [61]	94.42
	Xu *et al.* [62]	95.64

Tobacco-3482

Method	Model	Accuracy(%)
Image	Nicolas *et al.* [7]	84.5
Text		73.8
Multi-modal		87.8
Image	Asim *et al.* [6]	93.2
Text		87.1
Multi-modal		95.8
Image	Ferrando *et al.* [21]	94.04
Text		-
Multi-modal		94.90
Image	Cross-Modal [10]	96.25
Text		**97.18**
Multi-modal		**99.71**
Image	$EAML_{Tr-KLD_{Reg}}$ (Ours)	**97.99**
Text		96.27
Multi-modal		98.57
Baselines	Kumar *et al.* [34]	43.8
	Kang *et al.* [30]	65.37
	Afzal *et al.* [1]	76.6
	Harley *et al.* [25]	79.9
	Noce *et al.* [43]	79.8

Ablation Study: Inter-Dataset Evaluation Protocol

Inter-Dataset	Train	Test
	RVLCDIP	Tobacco-3482
	Tobacco-3482	RVLCDIP

Ablation Study :

Intra-Dataset & Inter-Dataset Evaluation Comparison

- **Train on Tobacco-3482 ▸ ▸ Test on Tobacco-3482**

Method	Image Modality	Text Modality	Fusion Modality
Positive Mutual Learning (**Ours**)	**97.70**	**96.27**	**98.28**
EAML	**97.99**	**96.27**	**98.57**

- **Train on RVLCDIP ▸ ▸ Test on Tobacco-3482**

Method	Image Modality	Text Modality	Fusion Modality
Positive Mutual Learning (**Ours**)	**84.82**	**83.73**	**86.68**
EAML	**87.29**	**87.23**	**87.63**

Ablation Study :

Intra-Dataset & Inter-Dataset Evaluation Comparison

- **Train on RVLCDIP ▸ ▸ Test on RVLCDIP**

Method	Image Modality	Text Modality	Fusion Modality
EAML	**97.67**	**97.63**	**97.70**

- **Train on Tobacco-3482 ▸ ▸ Test on RVLCDIP**

Method	Image Modality	Text Modality	Fusion Modality
EAML	**78.89**	**79.06**	**86.68**

D’ici, on voit + loin !

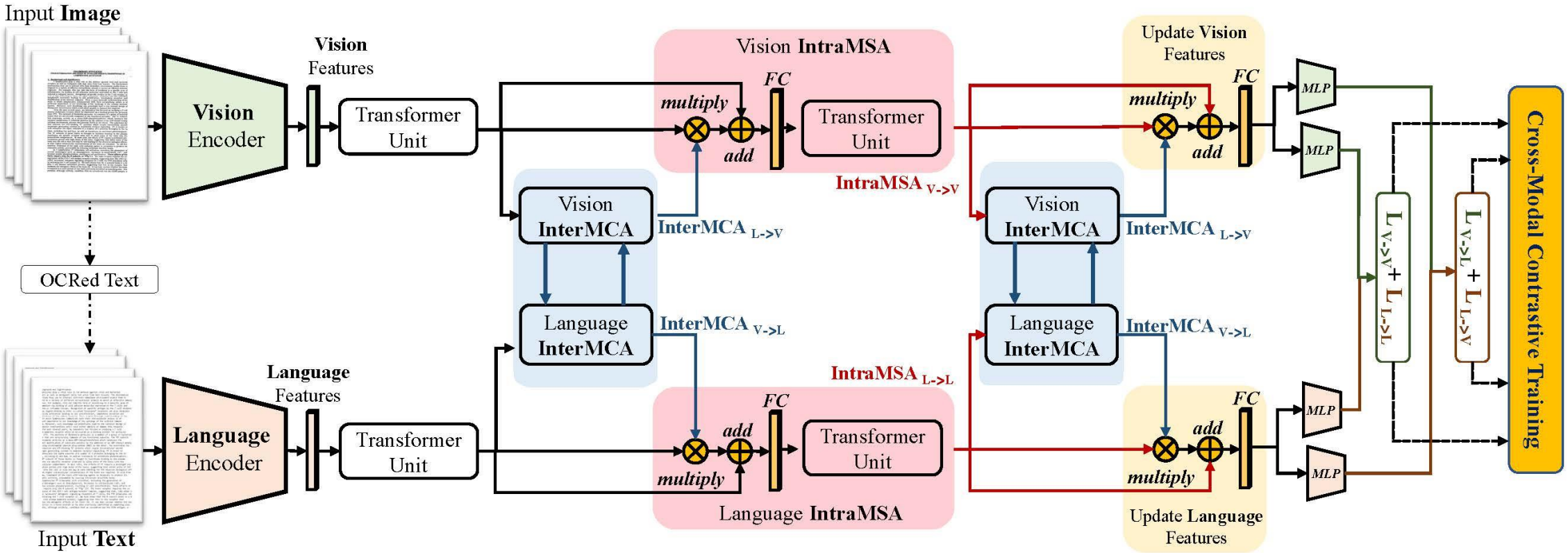
La Rochelle Université

univ-larochelle.fr

Current Work: Cross-Modal Learning (i.e. pre-training) to perform Document Classification (as a downstream task)

- We approach the document classification problem by learning cross-modal representations through language and vision cues, considering **intra- and inter-modality relationships.**
- Instead of merging features from different modalities into a common representation space, we intend to exploit high-level interactions and learn semantic information from effective attention flows **within and across** modalities.
- The learning objective that we propose aims to enforce the compactness of intra-class representations while separating inter-class features by contrasting positive and negative sample pairs within and across each modality.

Current Work: Cross-Modal Learning (i.e. pre-training) to perform Document Classification (as a downstream task)



+ *Overview of the proposed cross-modal contrastive learning method. The network is composed of InterMCA and IntraMSA modules with flexible attention mechanisms to learn cross-modal representations in a cross-modal contrastive learning fashion.*

Current Work: Cross-Modal Alignment

- We introduce the InterMCA and IntraMSA attention modules that capture intrinsic patterns by modeling the inter-modality and intra-modality relationships for image regions and texts.
- **Inter-Modal Alignment**:
  - The inter-modality cross-attention module **InterMCA** aims to enhance the cross-modal features by embracing cross-modal interactions across image regions and texts. This module aims to transfer the salient information from one modality to another.

$$\mathbf{InterMCA}_{\mathbf{L}\rightarrow\mathbf{V}}(\mathbf{V}^l) = \text{softmax}\left(\frac{\mathcal{Q}_{\mathbf{V}^l}\mathcal{K}^{\top}_{\mathbf{L}^l}}{\sqrt{d_k}}\right)\mathcal{V}_{\mathbf{L}^l} \qquad (1)$$

$$\mathbf{InterMCA}_{\mathbf{V}\rightarrow\mathbf{L}}(\mathbf{L}^l) = \text{softmax}\left(\frac{\mathcal{Q}_{\mathbf{L}^l}\mathcal{K}^{\top}_{\mathbf{V}^l}}{\sqrt{d_k}}\right)\mathcal{V}_{\mathbf{V}^l} \qquad (2)$$

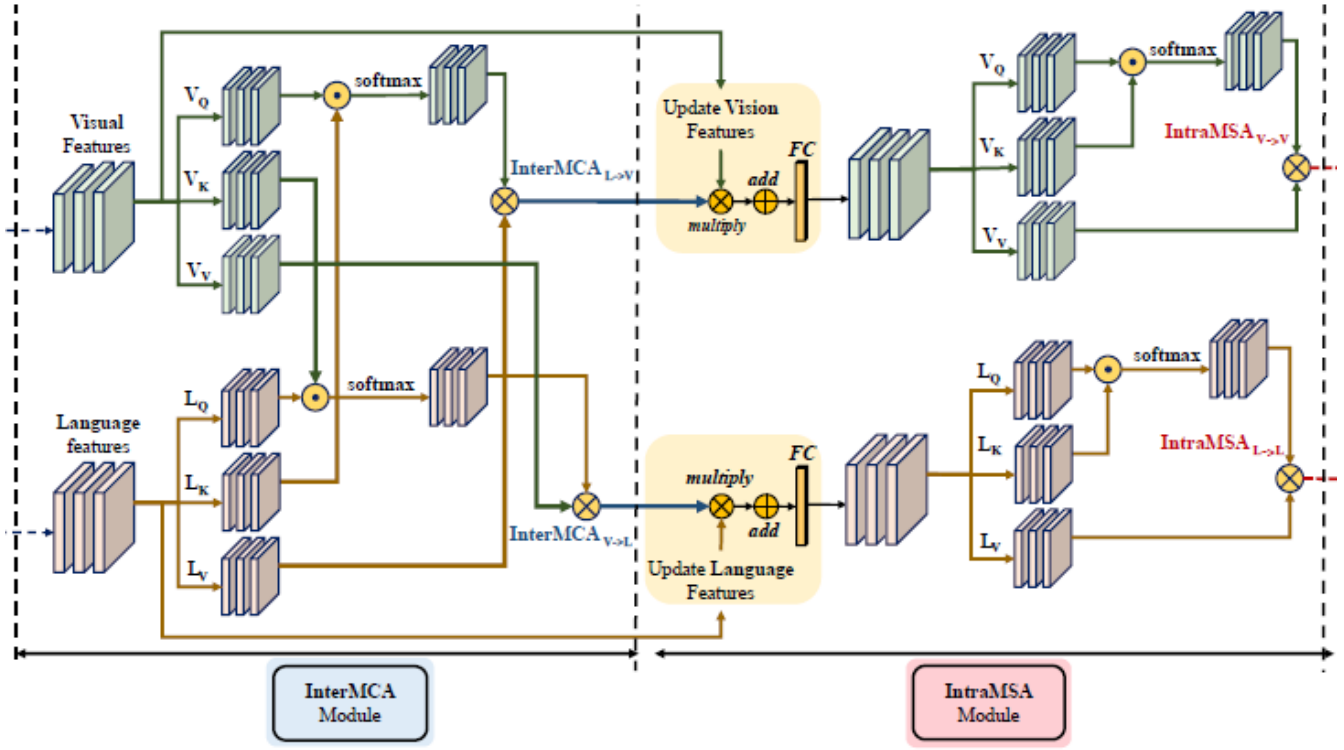
Current Work: Cross-Modal Alignment

- **Intra-Modal Alignment**:
  - The intra-modal self-attention module **IntraMSA** aims to update the vision and language information and to capture inner-modality attention weights. For each modality, the information is updated according to a feature fusion scheme.

$$\mathbf{IntraMSA}_{\mathbf{V}\rightarrow\mathbf{V}} = \mathrm{softmax}\left(\frac{\mathcal{Q}_{\hat{\mathbf{V}}^l}\mathcal{K}^{\top}_{\hat{\mathbf{V}}^l}}{\sqrt{d_k}}\right)\mathcal{V}_{\hat{\mathbf{V}}^l} \qquad (9)$$

$$\mathbf{IntraMSA}_{\mathbf{L}\rightarrow\mathbf{L}} = \mathrm{softmax}\left(\frac{\mathcal{Q}_{\hat{\mathbf{L}}^l}\mathcal{K}^{\top}_{\hat{\mathbf{L}}^l}}{\sqrt{d_k}}\right)\mathcal{V}_{\hat{\mathbf{L}}^l} \qquad (10)$$

Current Work: Cross-Modal Alignment

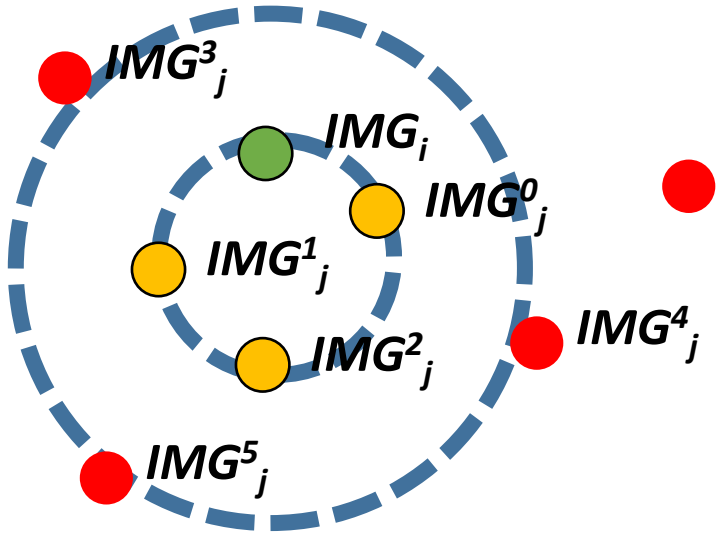


- *Illustration of the **InterMCA** and **IntraMSA** attention modules. The visual and textual features are transformed into query, key, and value vectors. They are jointly leveraged and are further fused to transfer attention flows between modalities to update the original features.*

Current Work: Cross-Modal Alignment

+ **Intra-Modality Learning Objective**

- *Positive Pairs for each Image are* ***(IMG_i, IMG_j)***

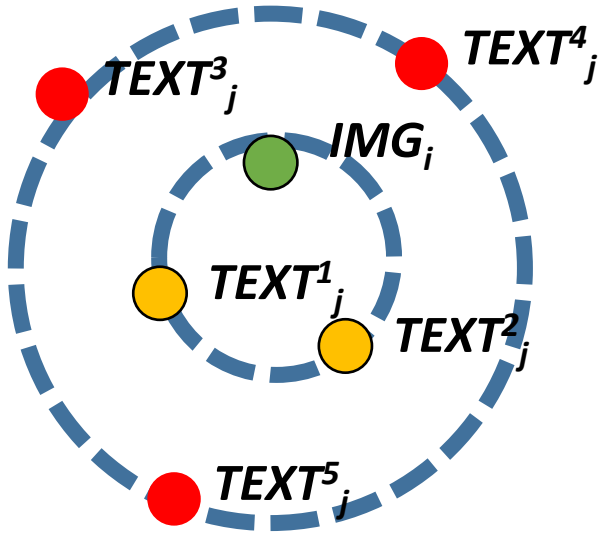


- *Negative Pairs for each Image are* ***(IMG_i, IMG_k)***

Current Work: Cross-Modal Alignment

+ **Inter-Modality Learning Objective**

- *Positive Pairs for each Image are* ***(IMG$_i$, TEXT$_j$)***



- *Negative Pairs for each Image are* ***(IMG$_i$, TEXT$_k$)***

Current Work: Results

Pre-training setting	IntraMSA	InterMCA	#Param	Acc.(%)
-w/o lang modality				
			198M	85.71
	√		201M	86.66
		√	209M	87.20
	√	√	217M	**90.94**
-w/o visn modality				
			198M	86.01
	√		201M	86.31
		√	209M	87.50
	√	√	217M	**90.62**

Table 1: Ablation study on CMDoC on cross-modality attention components, pre-trained on Tobacco dataset

Model	Accuracy (%)		
	T → R	R → T	R → N
w/o language modality			
- EAML [Bakkali *et al.*, 2021]	78.89	84.82	-
- **CMDoC**	**79.04**	**89.73**	**99.99**
w/o vision modality			
- EAML [Bakkali *et al.*, 2021]	79.06	83.72	-
- **CMDoC**	**81.96**	**89.88**	**99.99**

Table 3: Cross-dataset test on datasets with different size and document types. T,R, and N denote Tobacco, RVL-CDIP, and NIST datasets. T → R denotes fine-tune on Tobacco, and test on RVL-CDIP.

Current Work: Results

Method	Accuracy(%)
vision only methods	
VGG-16 [Afzal *et al.*, 2017]	90.31
AlexNet [Tensmeyer and Martinez, 2017]	90.94
Ensemble [Das *et al.*, 2018]	92.21
language only methods	
BERT_{Base} [Devlin *et al.*, 2019]	86.10
RoBERTa_{Base} [Liu *et al.*, 2019]	90.94
LayoutLM_{Base} [Xu *et al.*, 2020]	90.11
vision+language methods	
w/o language	
- Multimodal [Audebert *et al.*, 2019]	89.1
- Ensemble [Dauphinee *et al.*, 2019]	91.45
- EAML [Bakkali *et al.*, 2021]	90.81
w/o vision	
- Multimodal [Audebert *et al.*, 2019]	74.6
- Ensemble [Dauphinee *et al.*, 2019]	82.23
- EAML [Bakkali *et al.*, 2021]	88.80
CMDoC (V+L) w/o language	**92.64**
CMDoC (V+L) w/o vision	**91.37**
vision+language+layout methods	
TILT_{Base} [Powalski *et al.*, 2021]	93.50
SelfDoc [Li *et al.*, 2021]	93.81
LayoutLM_{Base} [Xu *et al.*, 2020]	94.42
LayoutLMv2_{Base} [Xu *et al.*, 2022]	95.25
DocFormer_{Base} [Appalaraju *et al.*, 2021]	96.17

Table 4: Top-1 accuracy (%) comparison results of different document classification methods evaluated on the of RVL-CDIP dataset. V+L denotes vision+language modalities

Future Work

- Essentially, we are exploring Self-Supervised Pre-training techniques to develop a more general and domain-agnostic multi-modal embedding network to be fine-tuned on downstream applications such as: few-shot document classification, document retrieval, domain generalization, etc.

References

Accepted Papers:

- S. Bakkali, Z. Ming, M. Coustaty and M. Rusiñol, "Cross-Modal Deep Networks For Document Image Classification," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 2556-2560, doi: 10.1109/ICIP40778.2020.9191268.

- S. Bakkali, Z. Ming, M. Coustaty and M. Rusiñol, "Visual and Textual Deep Feature Fusion for Document Image Classification," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 2394-2403, doi: 10.1109/CVPRW50498.2020.00289.

- Bakkali, S., Ming, Z., Coustaty, M. et al. EAML: ensemble self-attention-based mutual learning network for document image classification. IJDAR 24, 251–268 (2021). https://doi.org/10.1007/s10032-021-00378-0

In Progress:

- S. Bakkali, Z. Ming, M. Coustaty, M. Rusiñol, and OR. Terrades, "CMDoC: Cross-Modal Learning for Document Classification". (Under Review: Pattern Recognition Journal)

D’ici, on voit + loin !

La Rochelle Université

univ-larochelle.fr

La Rochelle Université

D'ici, on voit + loin !

EAML: Ensemble Self-Attention-based Mutual Learning Network for Document Image Classification

Souhail Bakkali, Mickaël Coustaty, Zuheng Ming, Marçal Rusiñol, Oriol Ramos Terrades

14/10/2022

La Rochelle Université

Informatique Image Interaction
Univ. La Rochelle

Centre de Visió per Computador

- PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR