



D'ici, on voit +loin !

Reconnaissance de textes et de personnages dans les bandes dessinées

Jean-Christophe BURIE

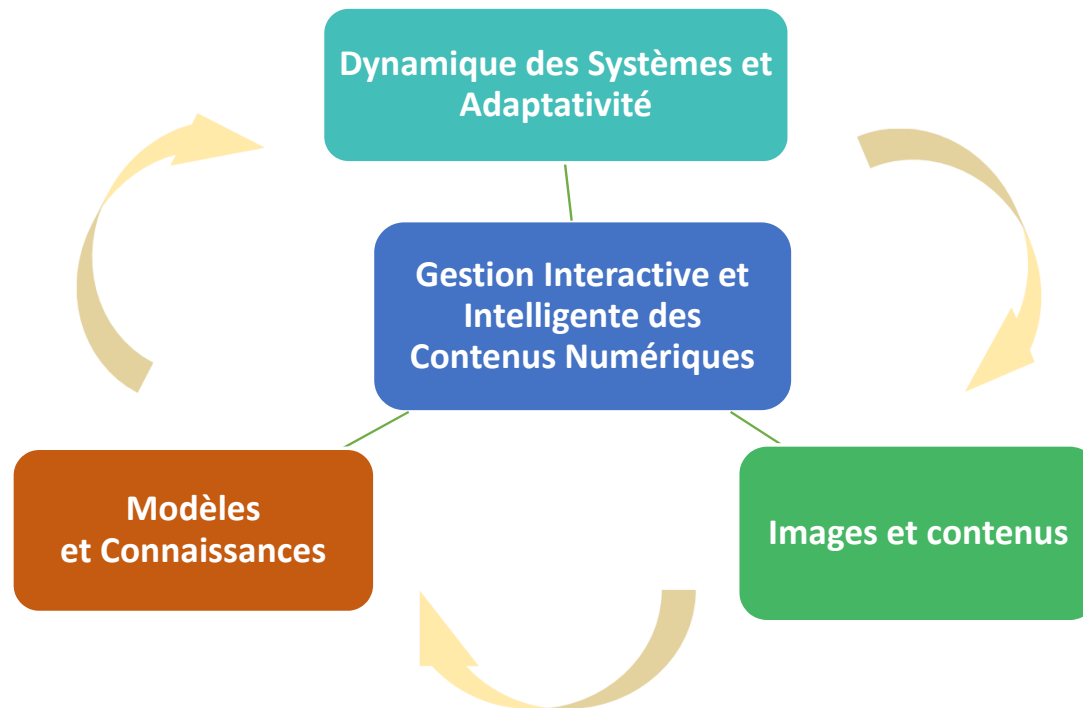
Laboratoire Informatique Image Interaction

- + Jean-Christophe Burie
Professeur en « Informatique »
- + Laboratoire Informatique Image Interaction
- + 32 permanents, 30 doctorants, 35 post-docs
- + Stratégie de recherche
*Gestion interactive et intelligente des
contenus numériques*



● La Rochelle

- + Trois équipes centrées sur la stratégie du laboratoire
 - > Volonté affirmée de décloisonnement scientifique entre les équipes



CONTEXTE

ED I2S : Module image 2019

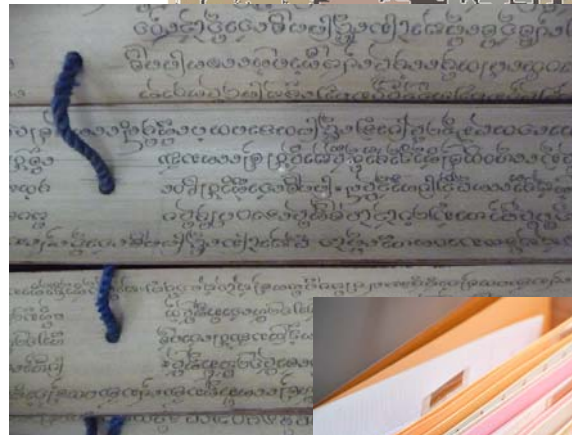
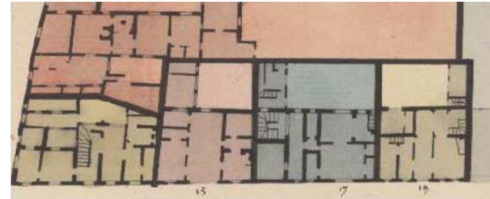
19 avril 2019 - Jean-Christophe Burie

Pourquoi la bande dessinée ?

+ Contexte scientifique > Equipe Images et Contenus

- ➔ Analyse de Documents
 - Plan de cadastre
 - Documents anciens
 - Lutte contre la fraude

- ➔ Dématérialisation
&
Indexation de contenus



Pourquoi la bande dessinée ?

+ Contexte territoriale

> Angoulême

- la **citô** internationale
de la bande dessinée
et de l'image

- Musée de la BD

- Festival de la BD



ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

- + Débuté en 2011
- + Objectifs :
 - > valoriser les bandes dessinées « **numérisées** » à l'aide des nouvelles technologies
 - > Proposer des approches **génériques** pour retrouver autant d'éléments que possible
- + Défis scientifiques
 - > Mélange de contenus graphiques et textuels
 - > Difficultés : document de formes libres et avec des contenus complexes
 - > Champ de recherche « récent » et « inexploré »
- + 2 thèses
 - > Extraction du contenu (thèse de Christophe Rigaud)
 - > Représentation des connaissances (thèse de Clément Guérin)

+ Extraction des éléments de base

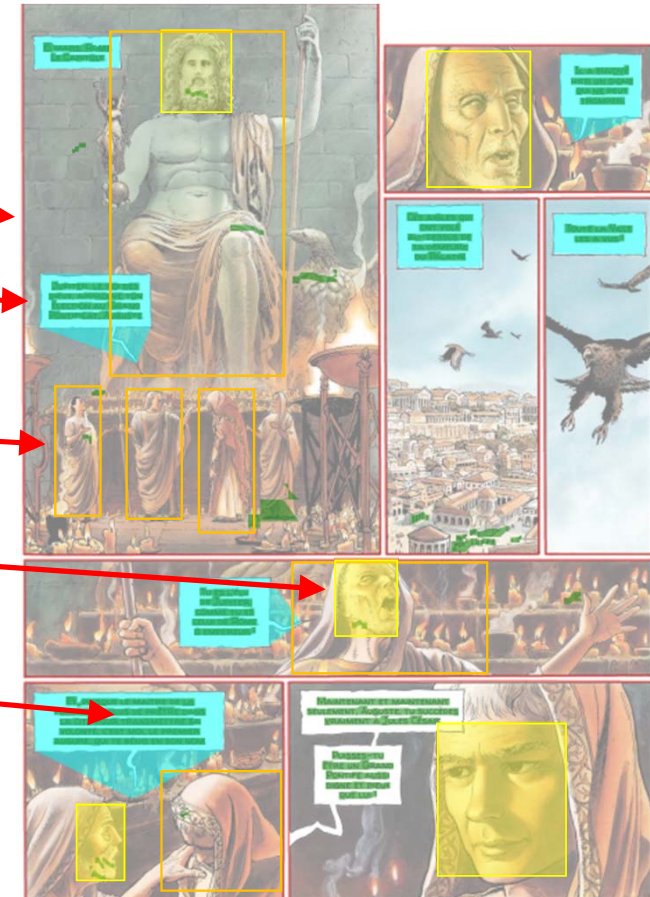
Cases

Bulles

Personnages

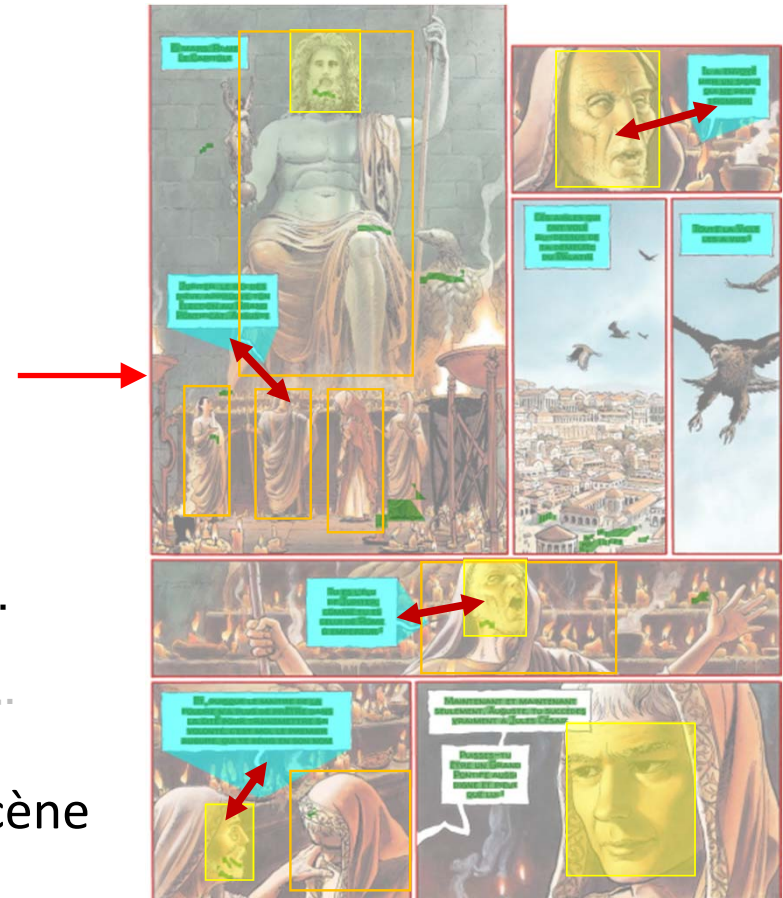
Visages

Texte



Objectif principal : Détection de la structure pour se focaliser sur l'analyse de chaque case

- + Extraction du contenu sémantique
 - > Reconnaissance du texte
 - ➔ Indexation plein texte
 - > Détecter l'ordre de lecture
 - > Lien entre les bulles et le personnage
 - ➔ Qui parle ? Que dit-il ?
 - > Reconnaissance des personnages
 - ➔ Qui est cet homme ? femme ? animal ? ...
 - > Reconnaissance des objets, lieu de l'action, ...



Objectif principal : Comprendre le contenu de la scène

+ Difficulté : grande variété de représentation

CASES



ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Difficulté : grande variété de représentation

BULLES

Image	<u>forme</u>	Contour
	<u>Ovale</u>	<u>Lisse</u>
	Rectangle	<u>Lisse</u>
	<u>Ovale</u>	<u>Ondulé</u>
	<u>Ovale</u>	<u>Hérissé</u>
	<u>Ovale / implicite</u>	<u>Lisse / Implicite</u>

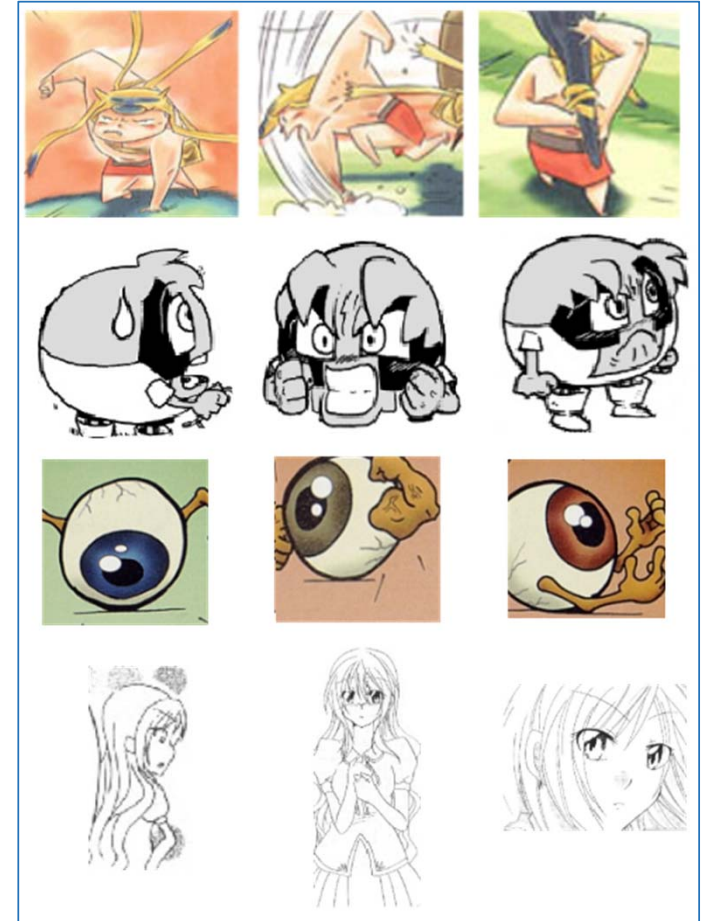
- + Difficulté : grande variété de représentation

TEXTE



- + Difficulté : grande variété de représentation

PERSONNAGES



- + Création d'une base de données pour évaluer les résultats
- + <http://ebdtheque.univ-lr.fr/>
- + Création d'un jeu de données hétérogènes Creation of heterogeneous dataset
 - > 100 pages de BD de 20 albums
 - > BD Franco-Belge "bandes dessinées", Comics américain, manga japonais
 - > de 1905 à 2012, papier and webcomics
 - > Obtention des droits pour usage scientifique auprès des éditeurs

- + Création d'une base de données pour évaluer les résultats
- + <http://ebdtheque.univ-lr.fr/>

Bibliographic annotations



PAGE (100)
Collection: Chilling Tales
Album: 17 Geo
Editor: Youthful Magazines
Drawer: Matt Fox
Writer: Matt Fox
Language: English
Page number: 16
Release date: 1953

Visual and semantic annotations



PANEL (850)
Rank: 1

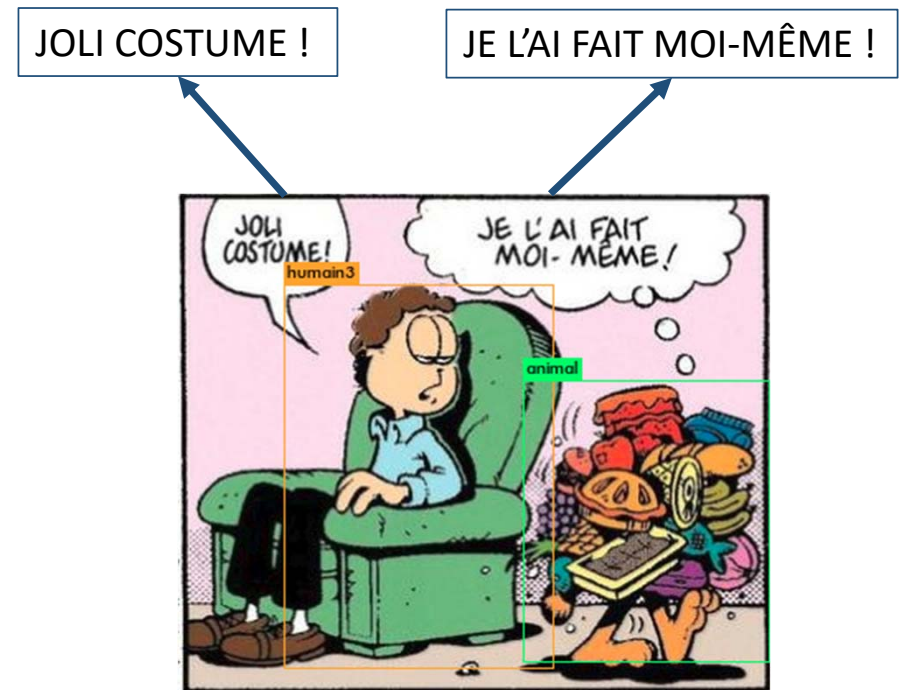
BALLOON (1092)
Rank: 2
Shape: Oval
Tail direction: South-West

TEXT LINE (4691)
Text: « STARK RAVING »

CHARACTER (1550)
LinkedToBalloon: 2

+ Comment

- > Détecter et reconnaître le texte ?
- > Détecter les personnages ?
- > ...



1 – CONTEXTE

2 – SEGMENTATION

+ **Segmentation et reconnaissance du texte**

+ **Reconnaissance des formes**

APPLICATION À LA BANDE DESSINÉE

3 – EXTRACTION ET RECONNAISSANCE DU TEXTE

4 – DÉTECTION DES PERSONNAGES

5 – CONCLUSIONS & PERSPECTIVES

SEGMENTATION

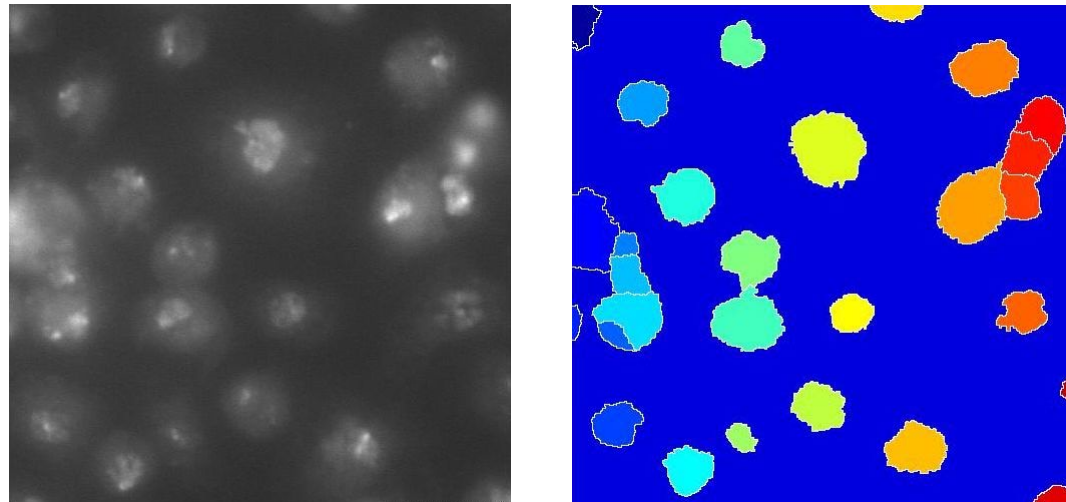
ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

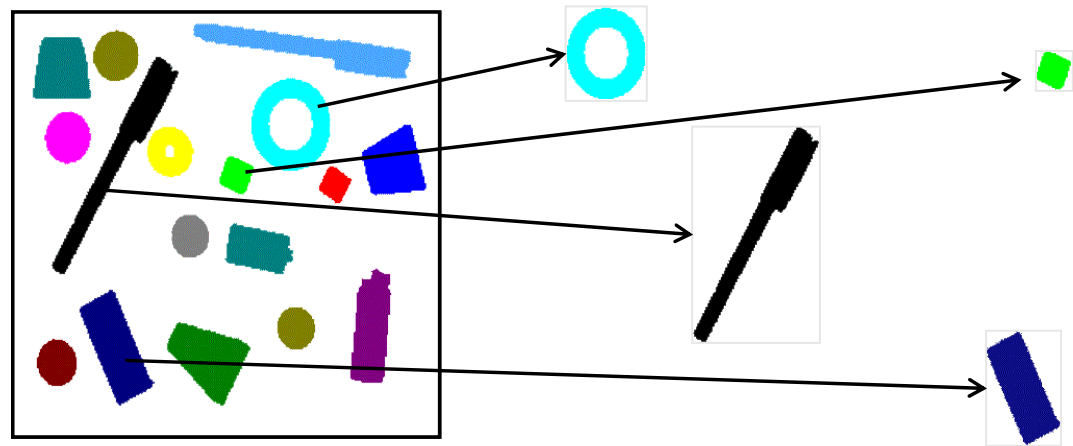
+ Segmenter

partitionner une image en zones homogènes selon un critère déterminé :

- > Niveau de gris
- > Couleur
- > Texture



- + Extraire (séparer) les entités d'une image pour
 - > appliquer un traitement spécifique à chacune
 - > interpréter le contenu de l'image (OCR, ICR, RdF)
- + Création d'un ensemble de composantes **connexes**



+ approche par seuillage

- > les regroupements de pixels sont faits en fonction de critères photométriques

+ approche « frontière »

- > les régions sont délimitées par les contours des objets qu'elles représentent (séparation)

+ approche « région »

- > les régions sont déterminées en fonction de leurs propriétés intrinsèques
(agrégation de pixels fonction d'un critère d'homogénéité)

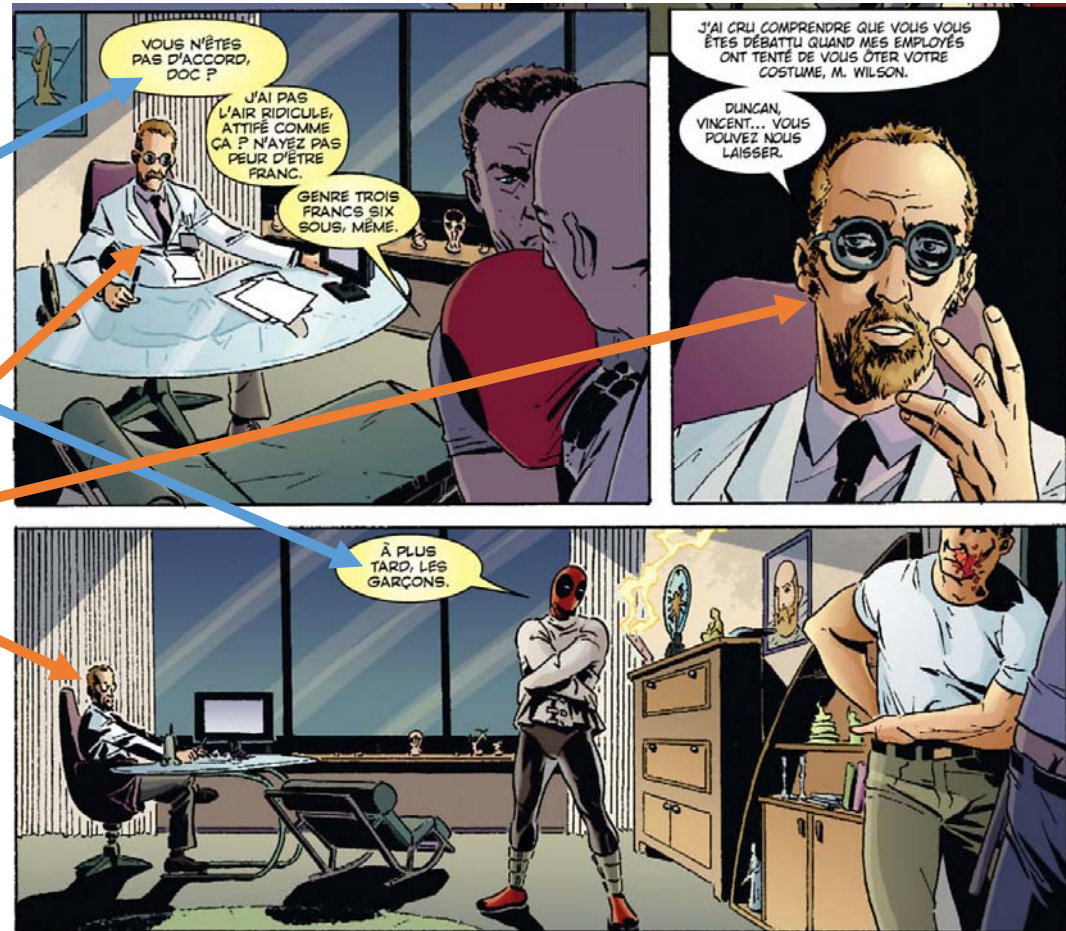
+ Apprentissage profond (deep learning)

- > Réseaux de neurones convolutionnels
- > Réseaux de neurones récurrents
- >

Cas de la Bande dessinée

Texte ?

Personnage ?



SEGMENTATION ET RECONNAISSANCE DU TEXTE

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

Segmentation du texte

- + Le texte est souvent très contrasté
 - > Noir sur fond blanc
- + Technique de seuillage efficace
 - > Seuil simple
 - > Seuil adaptatif
- + Plus difficile sur les documents anciens
 - > Dégradation
 - > Faible contraste
 - > transparence

BON DE COMMANDE

A RETOURNER :
MISTHAL FILM
(DVD)
65, rue de la Saunsière
92100 Boulogne-Billancourt

Au moment de la passation de commande, l'acheteur reconnaît avoir pris connaissance des conditions de vente décrites sur notre site et déclare expressément les accepter sans réserve.

VOS COORDONNEES :

Nom : _____ Prénom : _____
 Adresse : _____
 Code postal : _____ Ville : _____
 Tel : _____
 E-mail : _____

DATE :

Désignation	Prix Unitaire	Quantité	TOTAL TTC
(DVD) PNL "Le Conte du Monde Flammé et d'Autres Films d'Alain Couderc" d'une durée de 98 min	30,00 €		
TOTAL Hors frais de port			
Frais de Port (se reporter aux conditions de vente)			
TOTAL A PAYER			

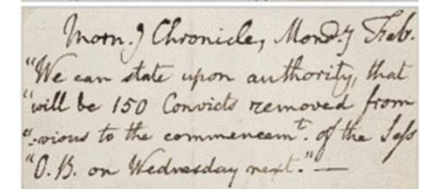
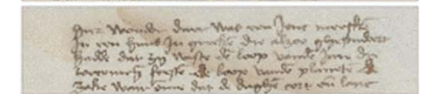
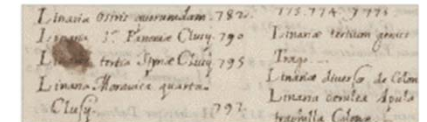
PAIEMENT :

☒ par chèque joint à l'ordre de Misthal Film
☐ par espèces.

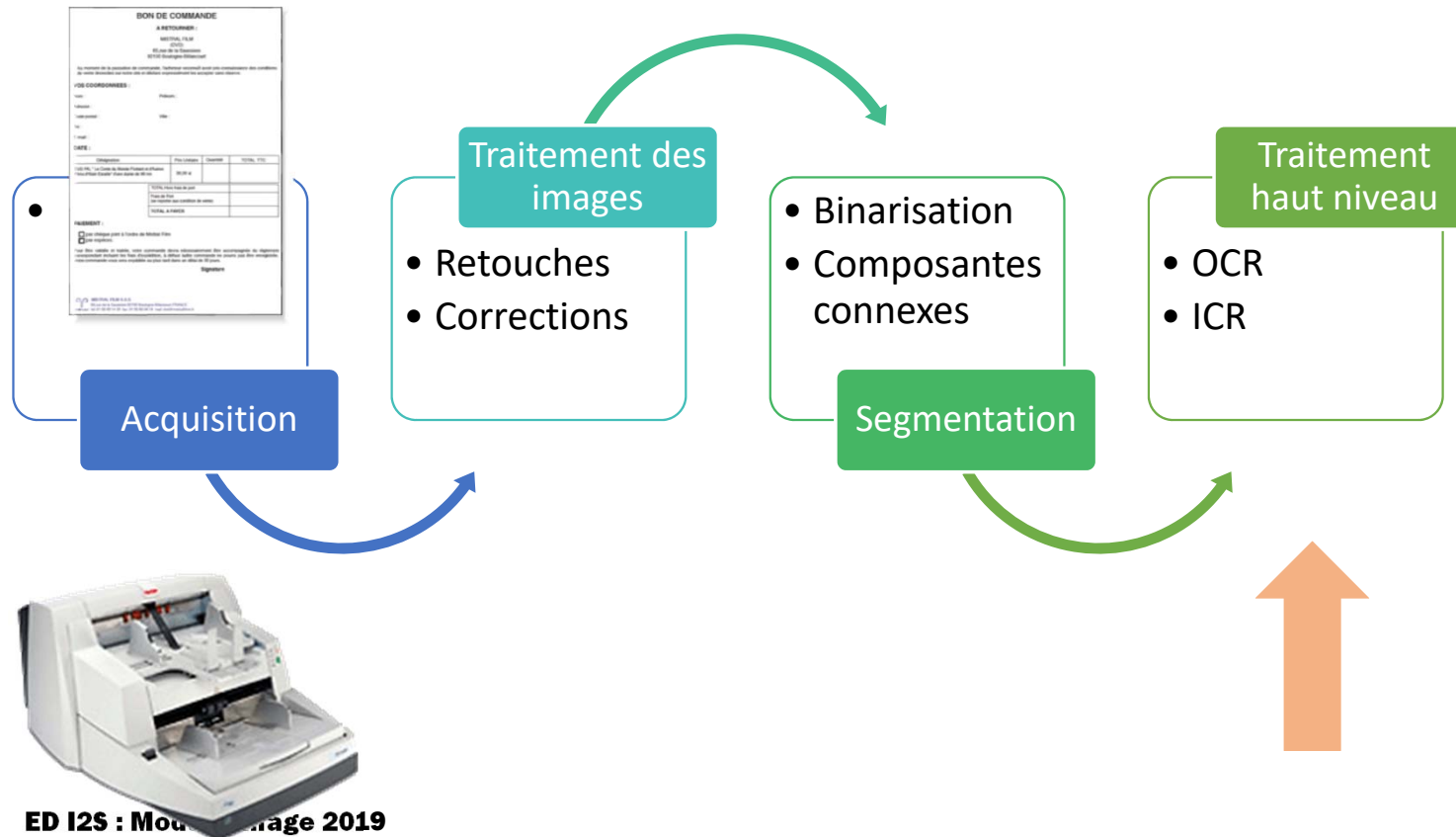
Pour être valable et traitée, votre commande devra nécessairement être accompagnée du règlement correspondant incluant les frais d'expédition, à défaut toutes commandes ne pouvant pas être enregistrées. Votre commande vous sera expédiée au plus tard dans un délai de 30 jours.

Signature _____

MISTHAL FILM S.A.S.
 65 rue de la Saunsière 92100 Boulogne-Billancourt FRANCE
 téléphone : 01 20 20 14 30 fax : 01 20 20 04 18 mail : misthal@misthal.fr



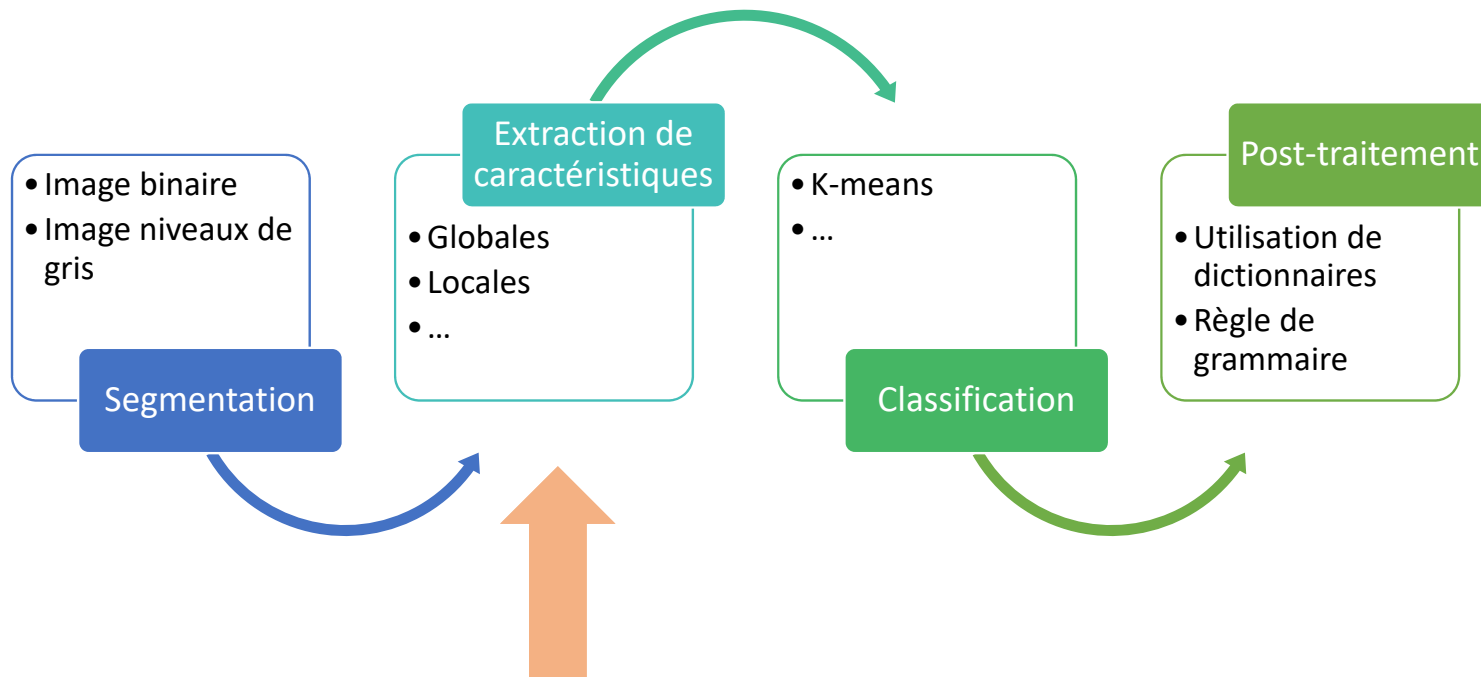
+ Chaîne de traitements d'un OCR/ICR



ED I2S : Module Image 2019

19 avril 2019 - Jean-Christophe Burie

+ Chaîne de traitements détaillée d'un OCR



+ D'une fonte à l'autre

- > Différents styles : graisse, stature (rapport hauteur/largeur), italique (inclinaison), contraste entre les pleins et déliés,
 - ➔ 10 classes de fontes différentes (Association typographique Int.)
- > Quelques valeurs :
 - En dactylographie : env 200 polices différentes (selon constructeur)
 - En imprimerie : catalogue d'un seul fournisseur (ex Monotype)
env 8000 polices avec tous les corps possible (échelle 6 à 100)
- ➔ Les logiciels OCR sont confrontés à des millions de formes possibles
(toutes variantes de fontes + variantes de qualité d'impression, ...)

+ Quelques exemples

ABCD h j k l 0 1 2 3 ! ') :

ABCD h j k l 0 1 2 3 ! ') :

ABCD h j k l 0 1 2 3 ! ') :

ABCD h j k l 0 1 2 3 ! ') :

ABCD h j k l 0 1 2 3 ! ') :

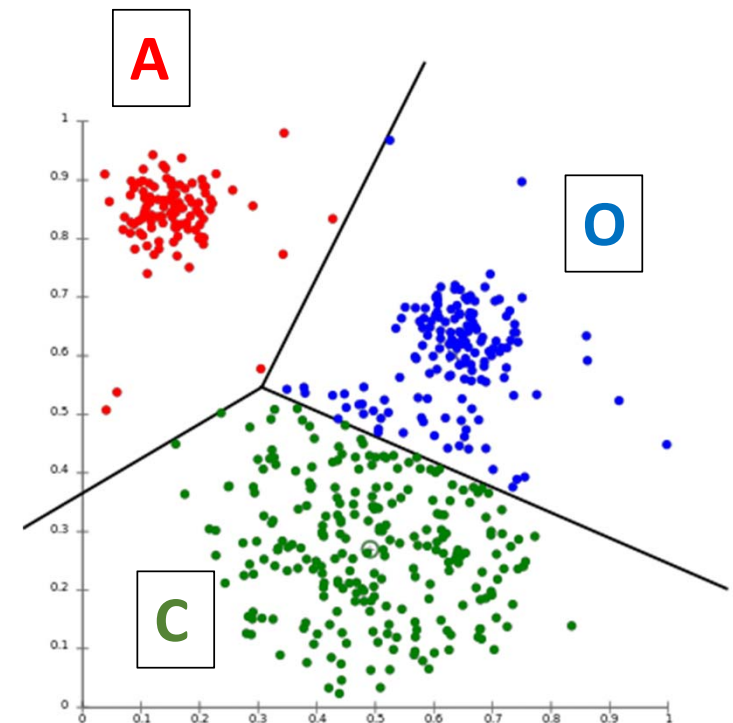
ABCD h j k l 0 1 2 3 ! ') :

ABCD h j k l 0 1 2 3 ! ') :

- + Une photocopie ou impression de mauvaise qualité engendre:
 - > Caractères collés, coupés ou du bruit de fond
- +
 - > Acquisition: problème d'échantillonnage, de numérisation, seuillage
- ➔ Accentue les phénomènes de déformation, collage ou cassure des caractères
- ➔ Nécessité d'extraire des caractéristiques robustes pour faire face à cette variabilité

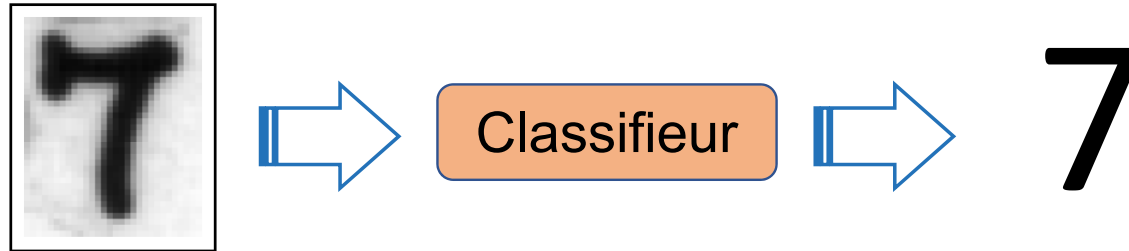
- + 1950-1970 : systèmes monofontes capables de reconnaître une seule fonte prédéfinie
- + 1980 - ... : systèmes multifontes à apprentissage (supervisé)
 - > Reconnaissance bas-niveau à partir de formes/modèles figées ou apprises.
- + 1990 - ... : systèmes capables utilisant des principes de reconnaissance symbolique
- + De nos jours : OCR classique = Extraction de caractéristiques (Features)

- + Caractéristiques = Features
- + Une forme représentée par un vecteur de valeurs numériques
- + Nombre de *features* est de l'ordre de 100 à 300.
- + Si les *features* sont bien choisies :
 - > Une classe de caractères (ex : A) sera représentée par un « nuage » contigu de points dans l'espace vectoriel des features.
 - > Rôle du classifieur = déterminer à quel nuage (donc à quelle classe de caractères) la forme à reconnaître appartient le plus vraisemblablement



Extraction des caractéristiques

- + **Objectif** : Extraction d'éléments pertinents à partir des caractères pour former un vecteur des caractéristiques
- + **Utilisation** : Le vecteur est utilisé par le classifieur pour reconnaître le caractère.



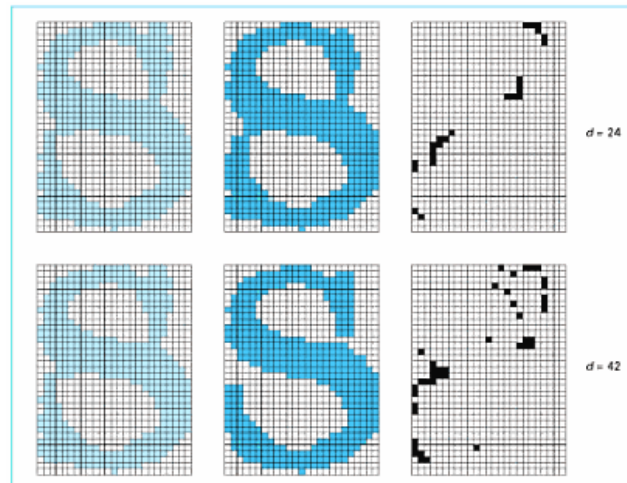
- + **Question** : Quelle méthode d'extraction de caractéristiques est la meilleure pour une application donnée ?

Quelles caractéristiques ?

- + De nombreuses caractéristiques peuvent être calculées à partir des caractères
- + Impossible de toutes les lister ...
- + Méthodes peuvent extraire les caractéristiques à partir :
 - > d'images en niveaux de gris
 - > des composantes connexes d'une image binaire
 - > du squelette de la forme
 - > des contours
- + Exemple de méthodes :
 - > Template matching, Templates déformables, Zoning, Projection d'histogramme, les moments (Géométrie, Hu, Zernike),

+ Template matching

- > Comparer le caractère à un modèle et trouver le plus ressemblant
- > Ressemblance estimée à partir d'un calcul de distance (comparaison pixel par pixel)



- > Approche limitée, nombreuses ambiguïtés (ex : S ou 8)

A partir d'une image en niveau de gris

+ Etape essentielle

- > Isoler les caractères dans l'image à niveaux de gris
 - soit directement à partir de l'image en niveaux de gris
(parfois compliqué sur les images difficiles)
 - soit à partir de l'image binaire
- > Chaque caractère = une imagerie

+ Moments invariants de HU

- > Proposé par M. K. Hu en 1962
- > Moments invariants à la translation, rotation et changement d'échelle
- > Les moments d'ordre $p+q$ sont définis par :

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

- > Les moments centrés sont définis par :

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy$$

$$\text{avec} \quad \bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

A partir d'une image en niveau de gris

+ Moments invariants de HU

> L'invariance au changement d'échelle est obtenu par normalisation :

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = (p+q+2)/2, \quad p+q=2,3;\dots$$

> Hu en déduit 7 moments normalisés

- très utilisés pour décrire les formes (les caractères)
- Invariant à la rotation ou au changement d'échelle

+ Moments invariants de HU

> 7 moments de Hu

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2$$

$$\begin{aligned} \phi_5 = & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned}$$

$$\begin{aligned} \phi_6 = & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \end{aligned}$$

$$\begin{aligned} \phi_7 = & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned}$$

+ Moments invariants de Zernike

- > Proposé par Zernike en 1934
- > Egalement utilisés sur les images binaires
- > Les moments d'ordre p et de répétition q sont définis par :

Où :

$$Z_{pq} = \frac{p+1}{\pi} \int_{\theta=0}^{2\pi} \int_{r=0}^1 V_{pq}(r, \theta) f(r, \theta) r dr d\theta$$

- $p = 0, 1, 2, \dots$ définit l'ordre.
- q est un entier vérifiant les contraintes : $0 \leq |q| \leq p$ et $p - |q|$ est pair
- $f(r, \theta)$ est la fonction image (le caractère)
- $V_{pq}(r, \theta)$ est le polynôme de Zernike, exprimé en coordonnées polaire, où (r, θ) sont définis sur le disque unité.

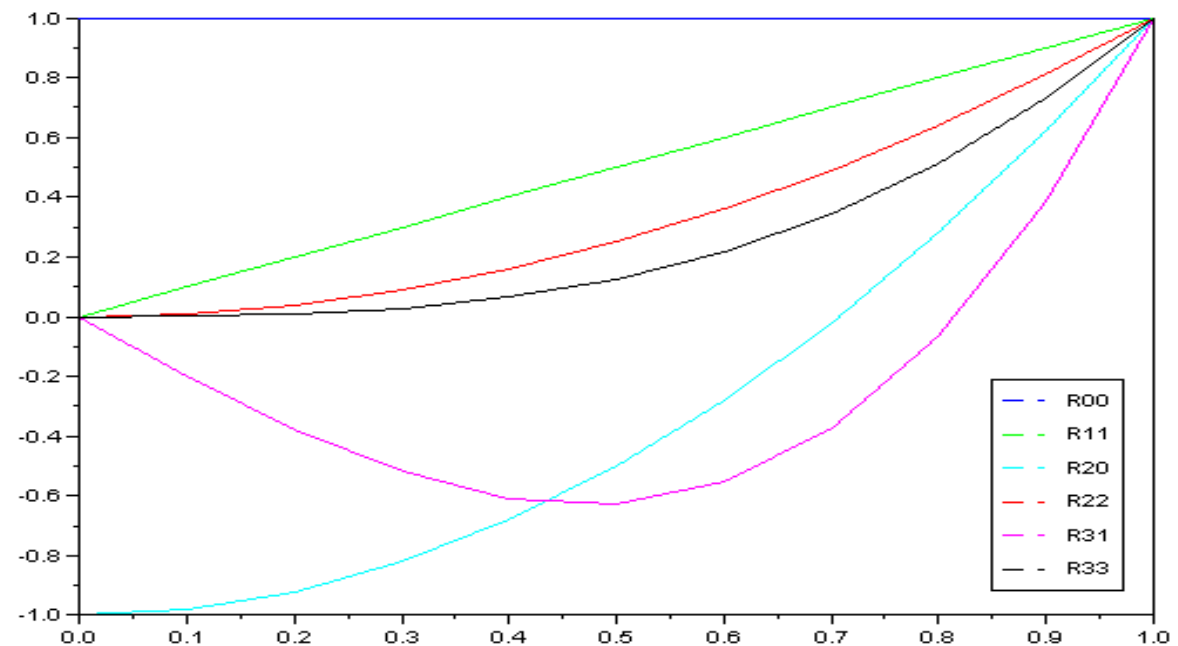
$$V_{pq}(r, \theta) = R_{pq}(r) e^{jq\theta}$$

- > $R_{pq}(r)$ est le polynôme radial orthogonal défini par :

$$R_{pq}(r) = \sum_{k=0}^{(p-|q|)/2} (-1)^k \frac{(p-k)!}{k!((p+|q|)/2-k)!((p-|q|)/2-k)!} r^{p-2k}$$

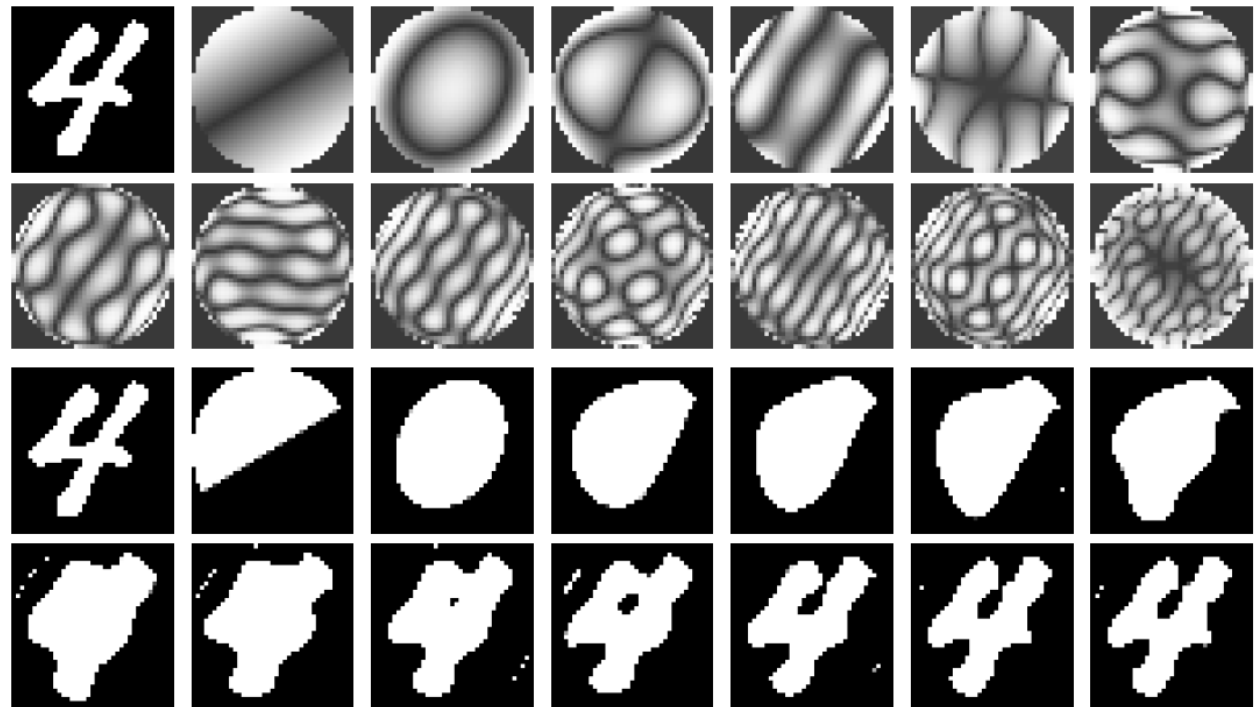
A partir d'une image en niveau de gris

- + Moments invariants de Zernike
 - > Exemple des premiers polynômes *radiaux*



A partir d'une image en niveau de gris

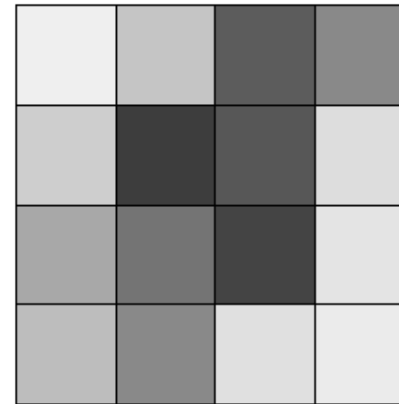
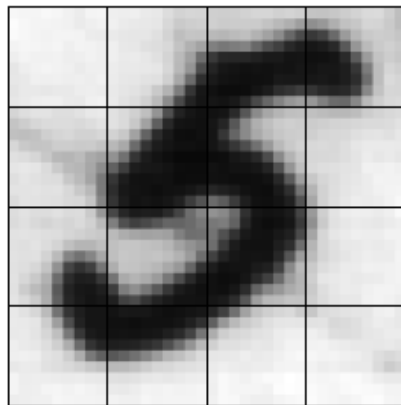
- + Moments invariants de Zernike
 - > Exemple : Moments d'ordre 1 à 13 pour le caractère 4.



A partir d'une image en niveau de gris

+ Grilles ajustables (Zoning)

- > Idée : découper l'image du caractère
- > Calculer l'intensité moyenne pour chaque zone

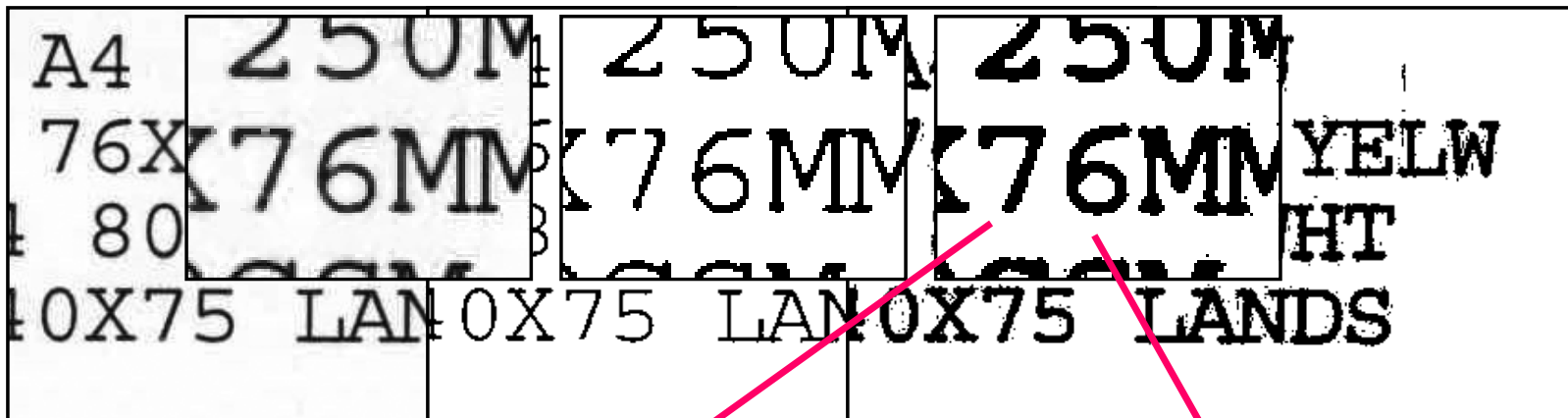


- > Si grille de $m \times n$ → vecteur de $m \times n$ intensités moyennes
- > Attention : pas d'invariance à la rotation, à l'illumination.

A partir d'une image binaire

+ Première étape

- > Isoler les caractères dans l'image binaire
 - Généralement : recherche des composantes connexes
 - Chaque composante connexe = caractère potentiel = une imagerie



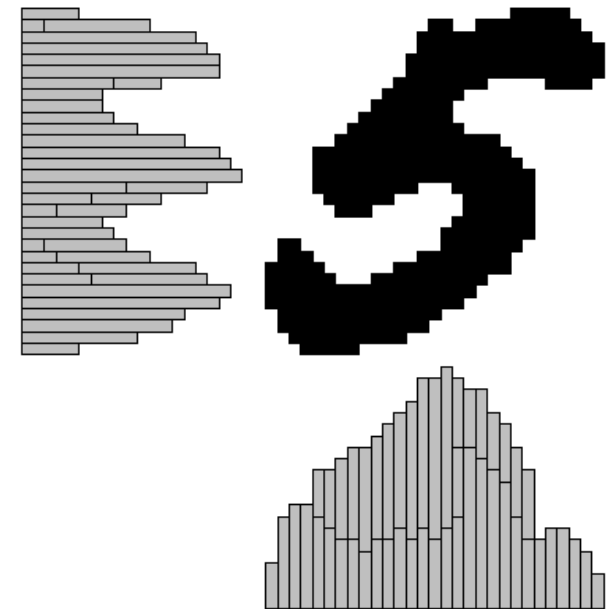
7

6

A partir d'une image binaire

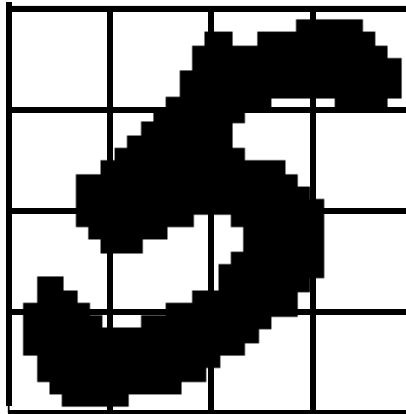
+ Histogrammes projetés

- > Comptabiliser le nombre de pixels horizontalement / verticalement
- > Pour être indépendant à l'échelle, utilisation :
 - d'un nombre de « cases » fixes
 - diviser les occurrences par le nombre de pixels
- > Les histogrammes servent de caractéristiques du caractère
- > Méthode sensible à la rotation
- > Possibilité de calculer également les histogrammes projetés sur les diagonales



A partir d'une image binaire

- + Grilles ajustables (Zoning)
 - > Idée : découper l'image du caractère
 - > Calculer le pourcentage de pixels noirs dans chaque zone



- > Si grille de $m \times n$ → vecteur de $m \times n$ pourcentage de pixels noirs

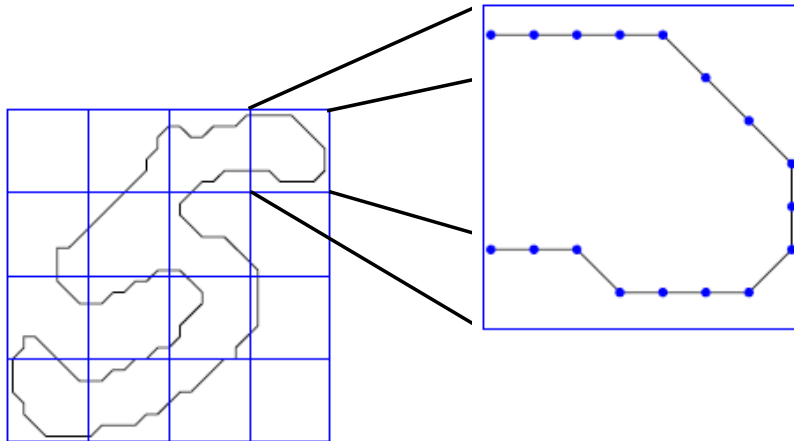
A partir des contours d'une image binaire

- + Première étape
 - > Extraire le contour des caractères
 - > Règle appliquée généralement
 - Contour extérieur = courbe passant par le centre de tous les pixels de contour 4-connexes au fond

A partir des contours d'une image binaire

+ Grilles ajustables (Zoning)

- > Idée : découper l'image du caractère
- > Calculer les orientations des segments de chaque case et les compatabiliser



Orientation	nb
0°	9
45°	1
90°	2
135°	4

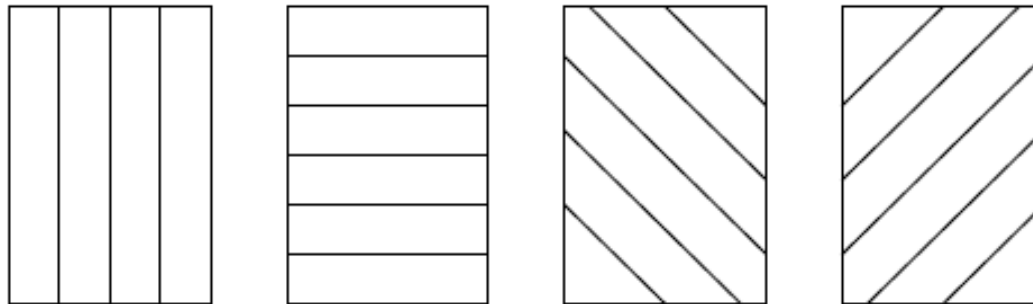
- > Si grille de $m \times n$ et k orientations → vecteur caractéristique de $m \times n \times k$ éléments

A partir des contours d'une image binaire

+ Grilles ajustables (Zoning)

> Autre approche :

réaliser les histogrammes des orientations selon le même principe que précédemment mais avec des zones horizontales, verticales et diagonales



A partir des contours d'une image binaire

+ Descripteurs elliptiques de Fourier

> Idée : modéliser le contour $(x(t), y(t))$ par les formules :

$$\hat{x}(t) = A_0 + \sum_{n=1}^N \left[a_n \cos \frac{2n\pi t}{T} + b_n \sin \frac{2n\pi t}{T} \right]$$

$$\hat{y}(t) = C_0 + \sum_{n=1}^N \left[c_n \cos \frac{2n\pi t}{T} + d_n \sin \frac{2n\pi t}{T} \right]$$

> Avec T la longueur du contour et n l'ordre

> les valeurs a_n, b_n, c_n, d_n sont utilisées comme caractéristiques

$$A_0 = \frac{1}{T} \int_0^T x(t) dt$$

$$C_0 = \frac{1}{T} \int_0^T y(t) dt$$

$$a_n = \frac{2}{T} \int_0^T x(t) \cos \frac{2n\pi t}{T} dt$$

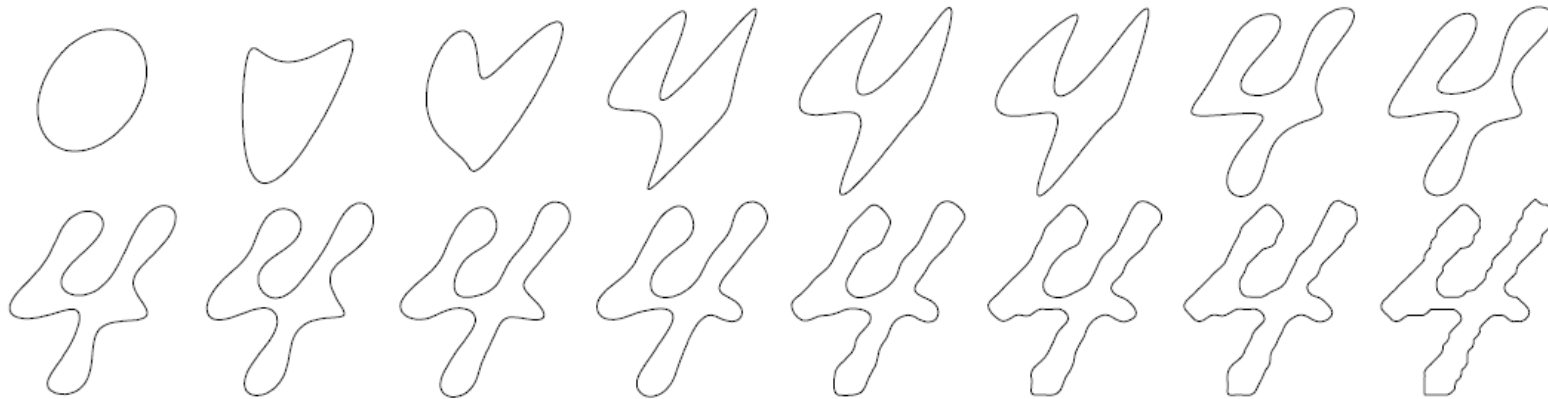
$$b_n = \frac{2}{T} \int_0^T x(t) \sin \frac{2n\pi t}{T} dt$$

$$c_n = \frac{2}{T} \int_0^T y(t) \cos \frac{2n\pi t}{T} dt$$

$$d_n = \frac{2}{T} \int_0^T y(t) \sin \frac{2n\pi t}{T} dt.$$

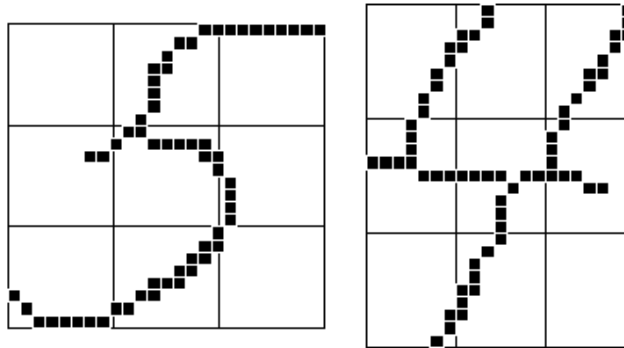
+ Descripteurs elliptiques de Fourier

- > Exemple : reconstruction du caractère '4' pour des descripteurs de Fourier d'ordre 1,2, ..., 10, 15, 20, 30, 40, 50, 100



+ Utilisation du squelette

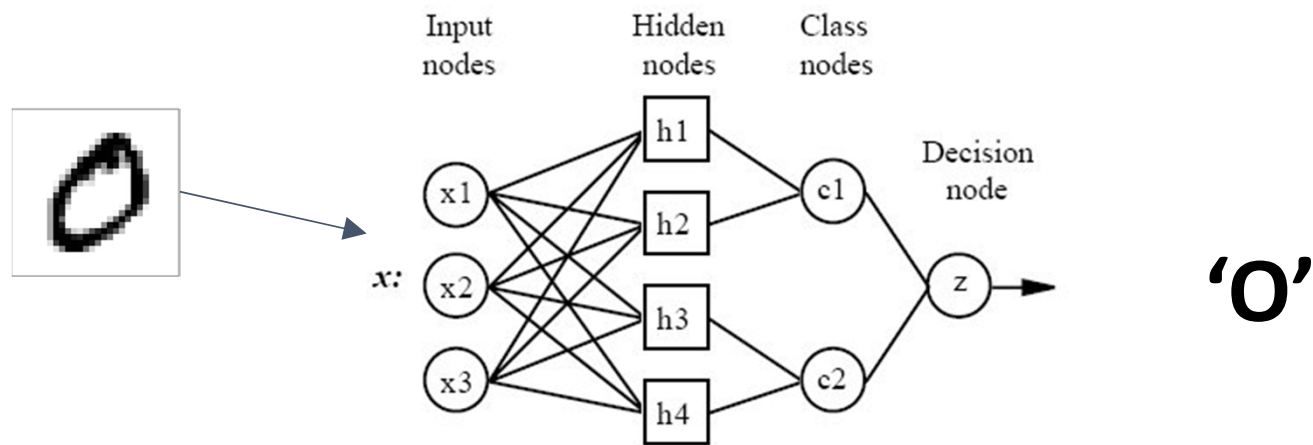
- > Consiste à affiner le caractère pour avoir un contour d'épaisseur 1 pixel
- > Grilles adaptatives
 - longueur du squelette dans chaque zone (normalisé en divisant par le nombre de pixels)



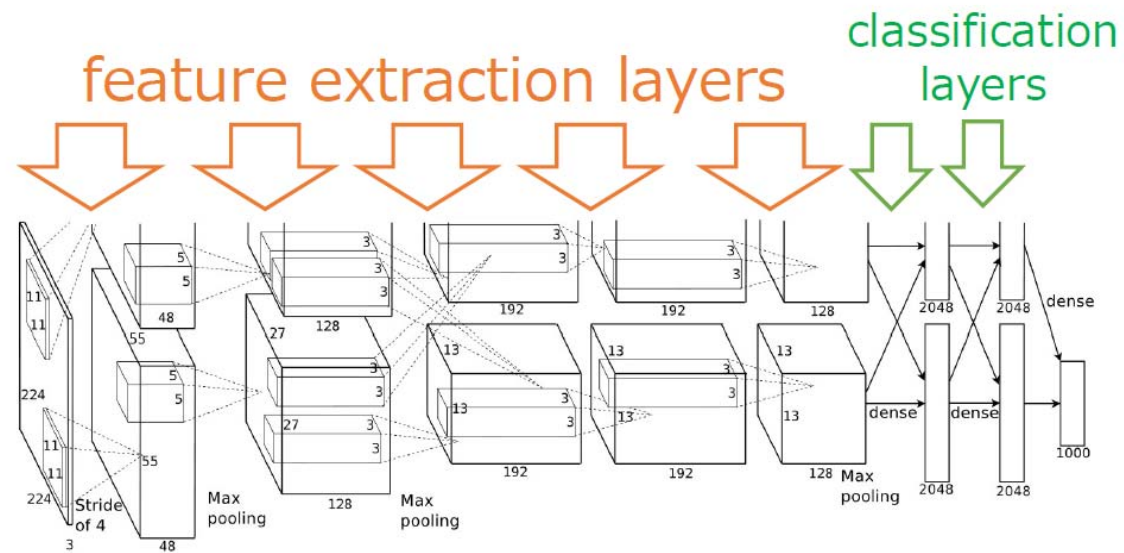
- + Sans oublier les descripteurs géométriques
 - > Certains descripteurs géométriques élémentaires permettent de réaliser un préclassement des images de caractères.
 - > Par exemple :
 - L'élongation : rapport longueur / hauteur
 - La compacité : rapport surface de l'objet sur carré de son périmètre
 - Le taux de remplissage : surface de l'objet sur surface de la boîte englobante
 - etc.

- + Il existe de nombreuses méthodes pour extraire les caractéristiques
- + Possibilités de combiner plusieurs méthodes
- + vecteurs de grande dimension pour caractériser chaque caractère
- + Le choix des caractéristiques dépend du classifieur utilisé.

+ Réseaux Classiques



- + Réseaux profonds
 - > CNN



+ Réseaux profonds

> Résultats

machine printed									
0	1	2	3	4	5	6	7	8	9
0	1	2	99.99%				7	8	9
0	1	2	3	4	5	6	7	8	9

handwritten									
0	1	2	3	4	5	6	7	8	9
0	1	2	99.79%				7	8	9
0	1	2	3	4	5	6	7	8	9

designed fonts									
0	1	2	3	4	5	6	7	8	9
0	1	2	95.49%				7	8	9
0	1	2	3	4	5	6	7	8	9



Proche des performances humaines

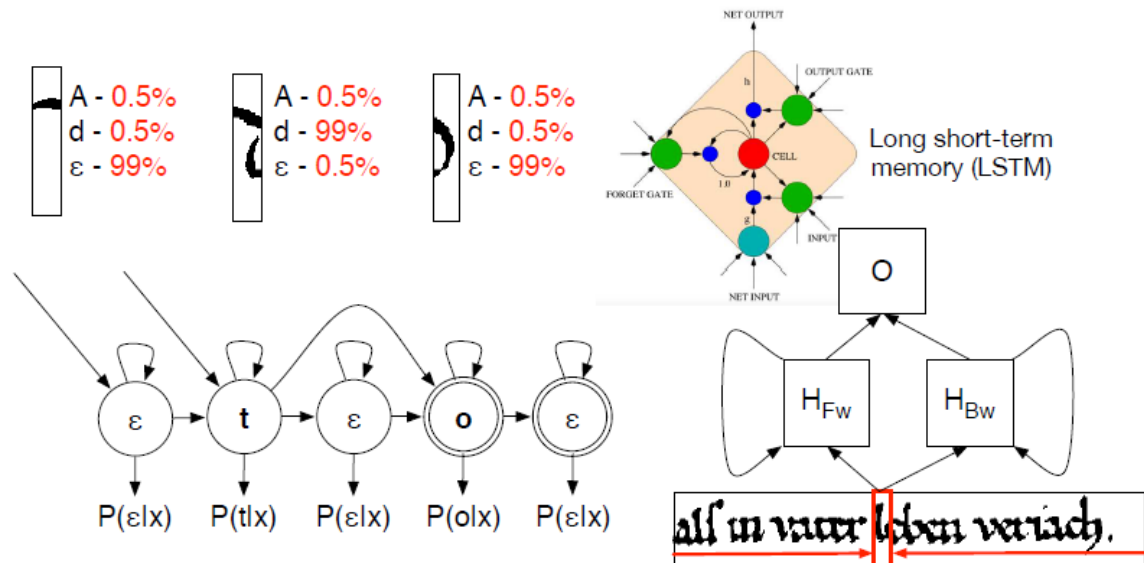
+ Réseaux de neurones récurrents

> Analyse Séquentielle

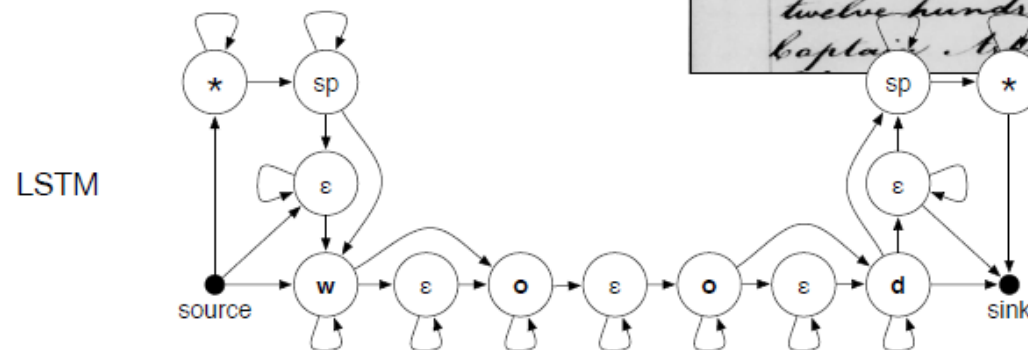
- Ligne de texte
- Texte manuscrit

+ LSTM

+ Nécessite apprentissage



- + Application à la recherche de mots
 - > Word spotting
- + Apprendre aux réseaux à reconnaître des mots clés



270. Letters, Orders and Instructions, October 1755.
only for the public use, unless by particular Orders from me. You are to send down a Barrel of Hints with the Arms, to Winchester, and about two thousand weight of Flour, for the two Companies of Rangers; twelve hundred of which to be delivered Captain Abby and Company, at the

RECONNAISSANCE DES FORMES

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Retrouver des motifs



+ La reconnaissance de motifs peut être effectuée au moyen de divers algorithmes d'apprentissage automatique tels :

- > réseau de neurones ;
- > une analyse statistique ;
- > l'utilisation de modèles de Markov cachés ;
- > une recherche d'isomorphisme de graphes ou de sous-graphes.
- > ...

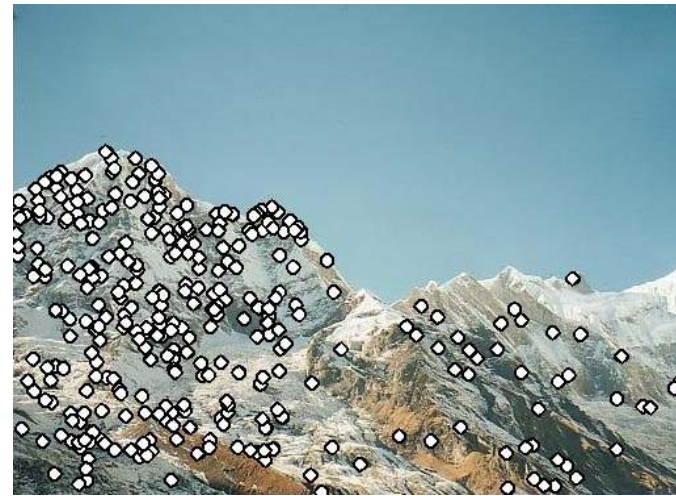
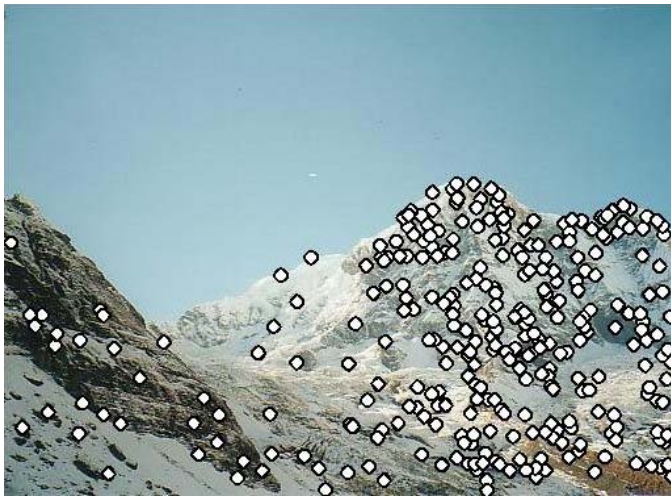
+ Exemple d'images naturelles



Cas des images complexes

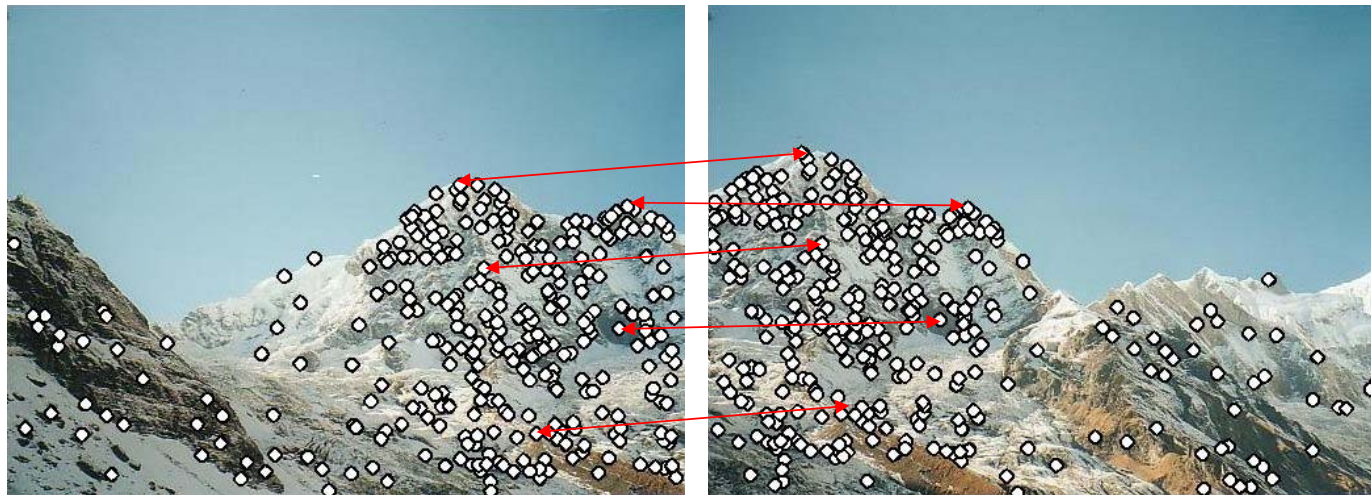
+ Détecter des points caractéristiques dans chaque image

- > Images naturelles
- > Images de documents
- > ...



Cas des images complexes

- + Détecter des points caractéristiques dans chaque image
- + Trouver les paires correspondantes

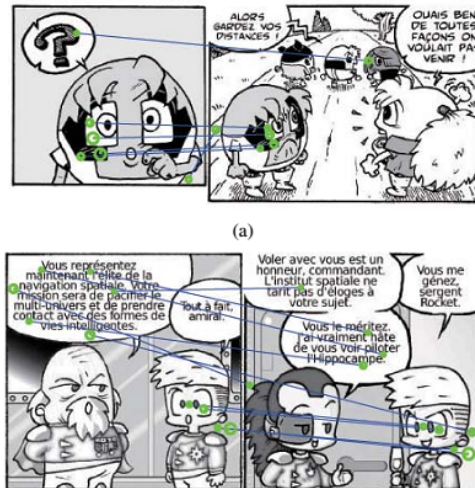


- + Détecter des points caractéristiques dans chaque image
- + Trouver les paires correspondantes
- + Opérations spécifiques
 - > Transformations géométriques (panorama)
 - > Comparaisons (Reconnaissance de formes)



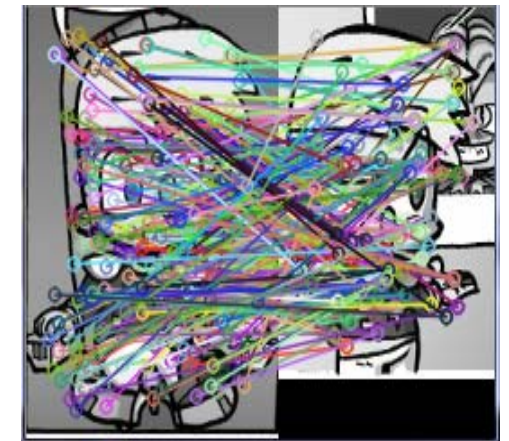
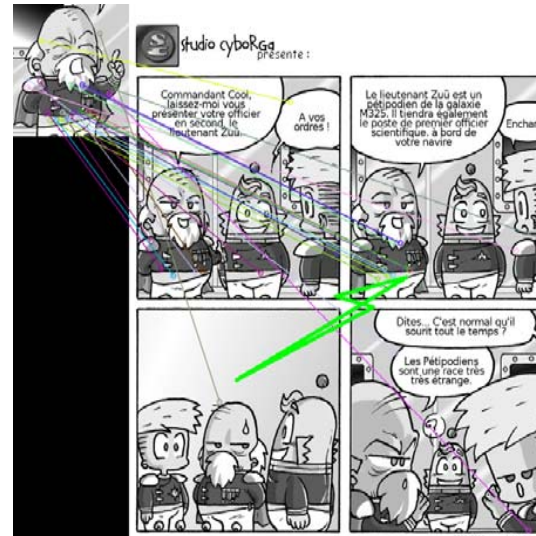
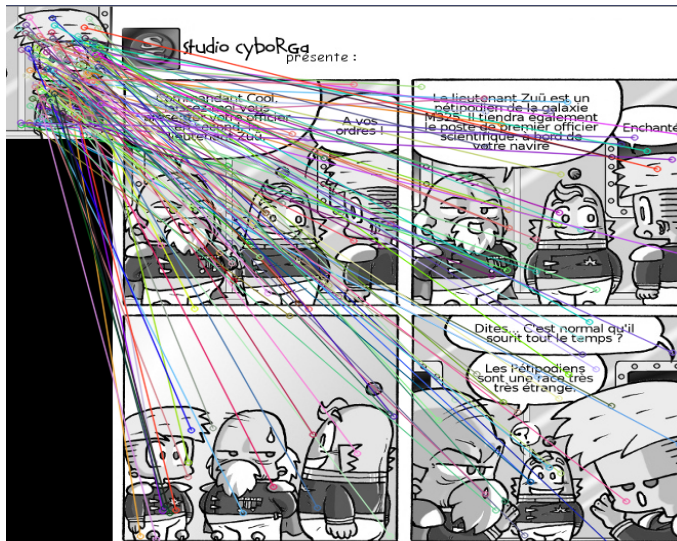
Exemple de détecteurs classiques

- + Détecteur de coins de Harris
- + Détecteur SIFT (Scale invariant Features Transform)
- + Détecteur SURF (Speeded Up Robust Features)



Exemple de détecteurs classiques

- + Détecteur de coins de Harris
- + Détecteur SIFT (Scale invariant Features Transform)
- + Détecteur SURF (Speeded Up Robust Features)

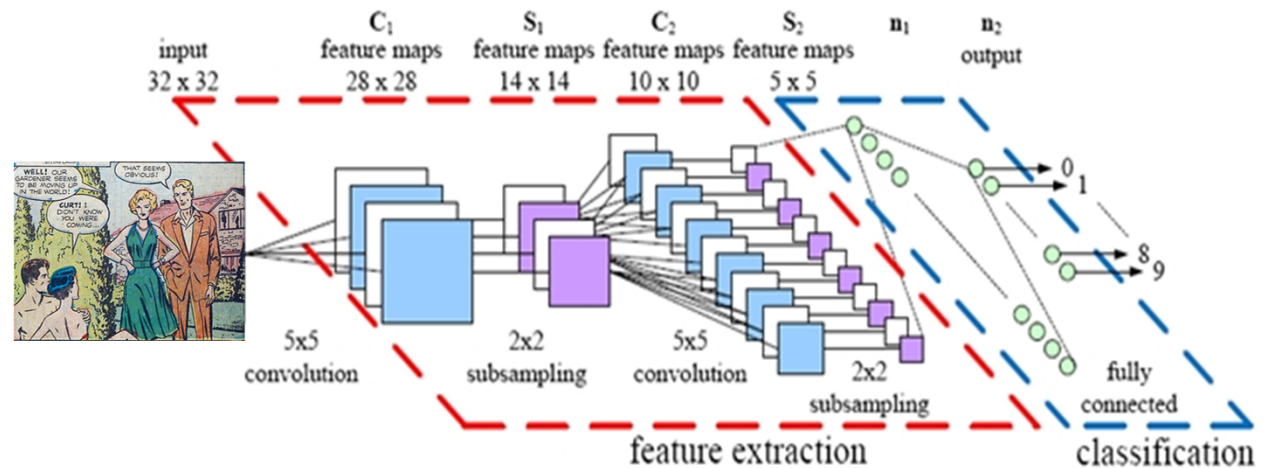


ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

- + Réseaux profonds
 - > CNN, ...

- + Nombreuses architectures
 - > FCN
 - > Faster-RCNN
 - > YOLO
 - > ...



1 – CONTEXTE

2 – SEGMENTATION

+ Segmentation et reconnaissance du texte

+ Reconnaissance des formes

APPLICATION À LA BANDE DESSINÉE

3 – EXTRACTION ET RECONNAISSANCE DU TEXTE

4 – DÉTECTION DES PERSONNAGES

5 – CONCLUSIONS & PERSPECTIVES



EXTRACTION ET RECONNAISSANCE DU TEXTE

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

EXTRACTION DU TEXTE

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

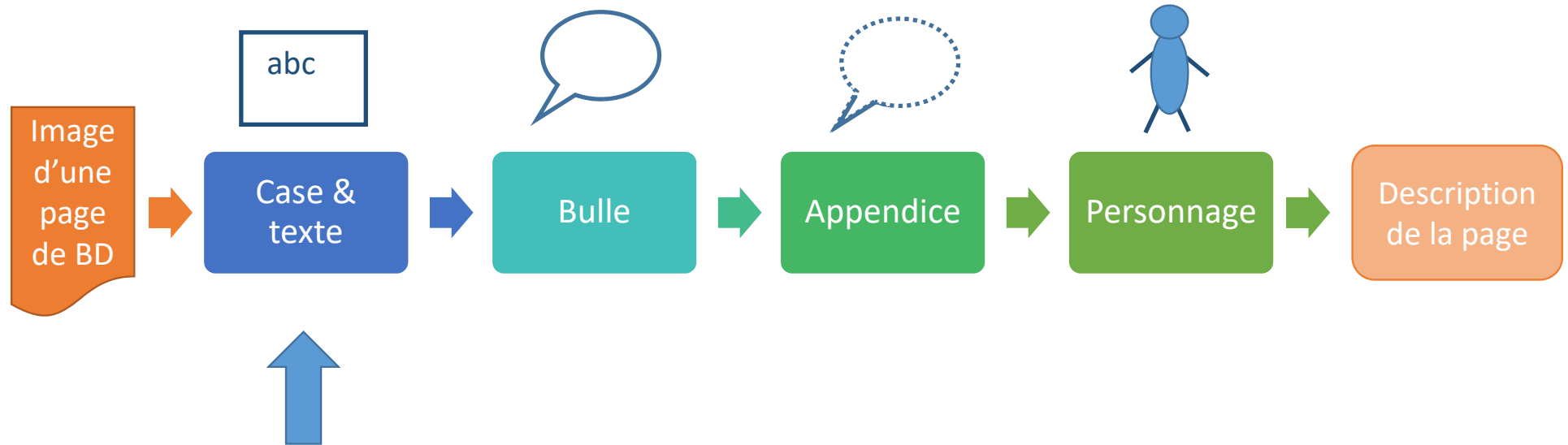
Première approche



ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Analyse séquentielle



Extraction du texte (et des cases)

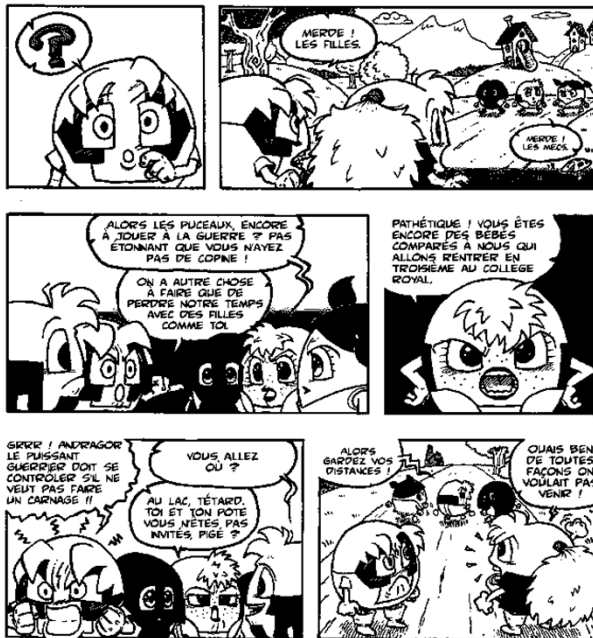
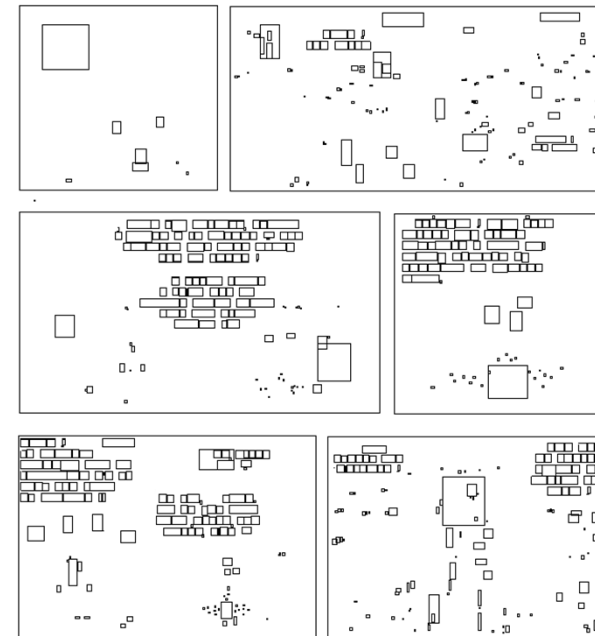


Image binaire

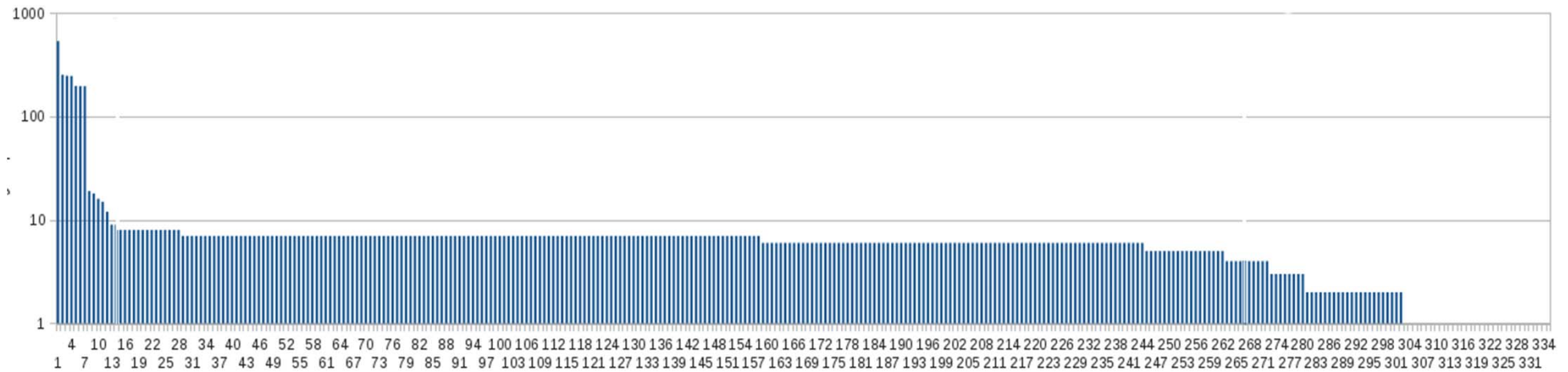


Boites englobantes des composantes connexes (CC)

Extraction du texte (et des cases)

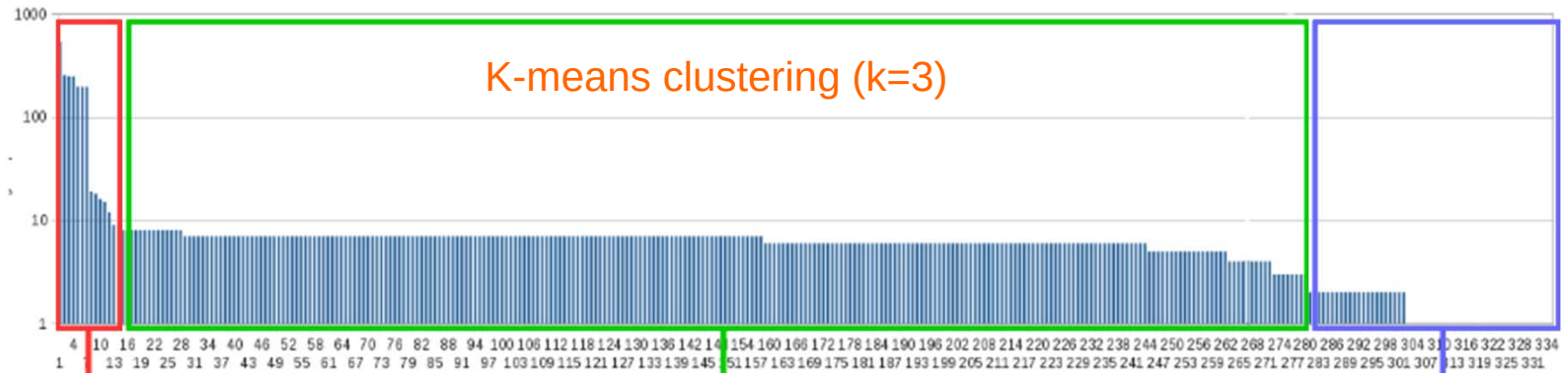
Histogramme des composantes connexes par ordre décroissant

Hauteur
des CC
(pixel)

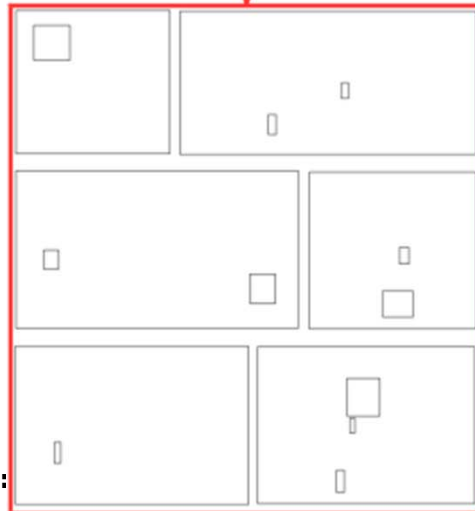


Extraction du texte (et des cases)

Hauteur des CC
(pixel)



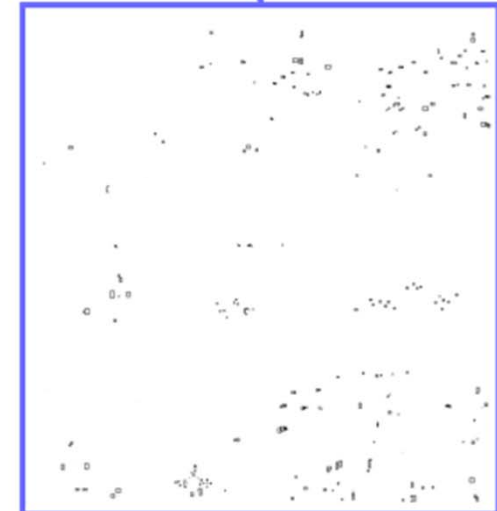
CASE



TEXTE



BRUIT

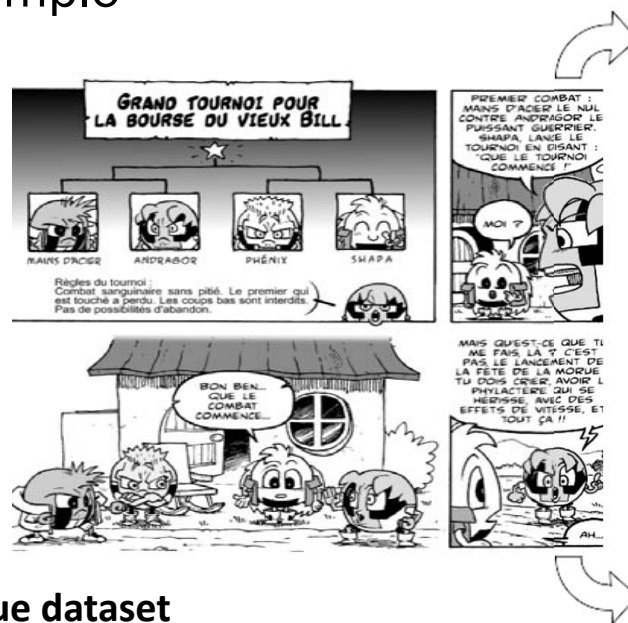


ED I2S :

19 avril 2019 - Jean-Christophe Burie

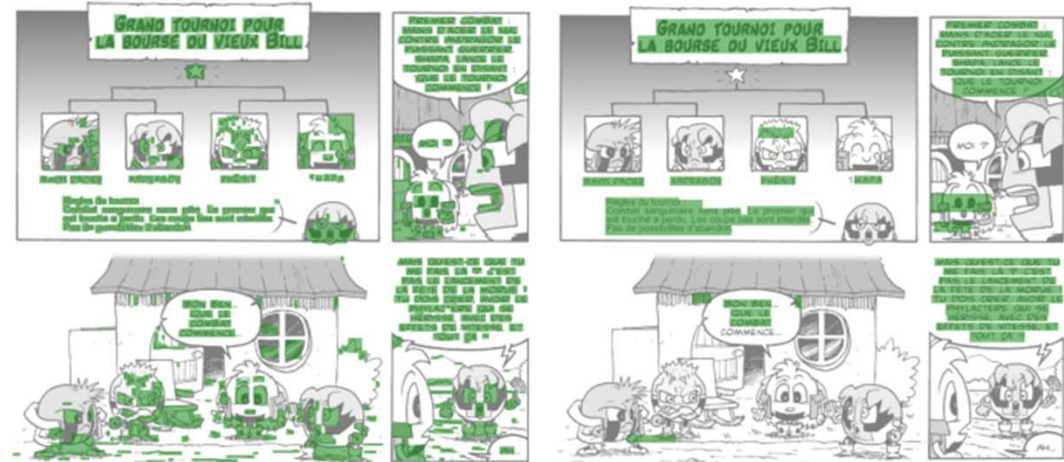
Extraction du texte (et des cases)

+ Exemple



Grandes CC

Cases



Petites CC

Texte

eBDtheque dataset

	Rappel	Précision
Cases	64.24 %	83.81 %
Texte	54.91 %	57.15 %

ED I2S : Module image 2019

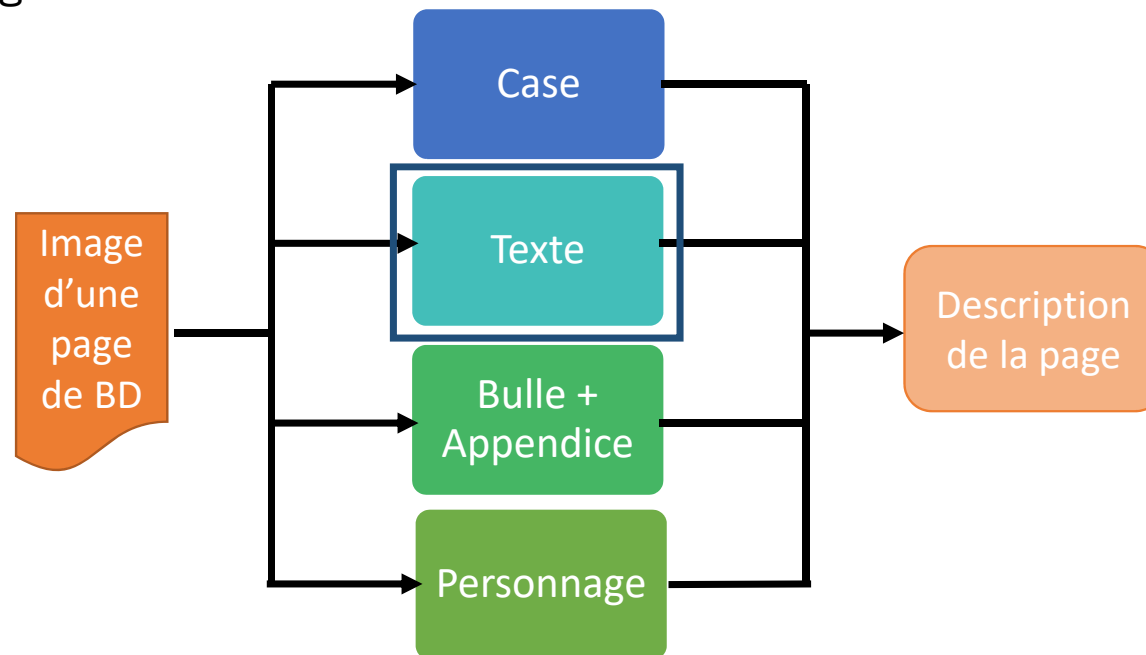
19 avril 2019 - Jean-Christophe Burie



ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

- + Approche parallèle
 - > Eviter la propagation d'erreur



+ Principe

- > Les zones de texte sont les régions les plus contrastées (noir sur blanc)

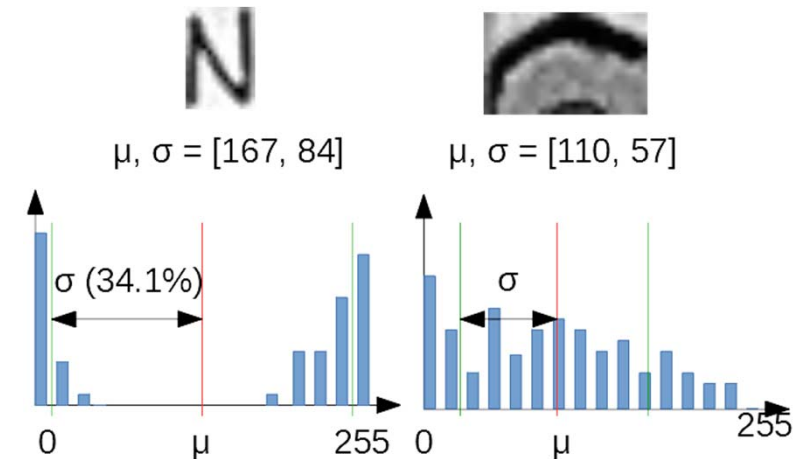
+ Recherche des zones candidates

- > Comparaison de la distribution des niveaux de gris de chaque boite englobante des CC avec le contraste de la page
- > Classification des boites englobantes en fonction du contraste en utilisant la distance de Mahalanobis

Contraste de la page $c = I_{\max} - I_{\min}$
Distance de Mahalanobis optimale $dm = 2$ (expérimentalement)

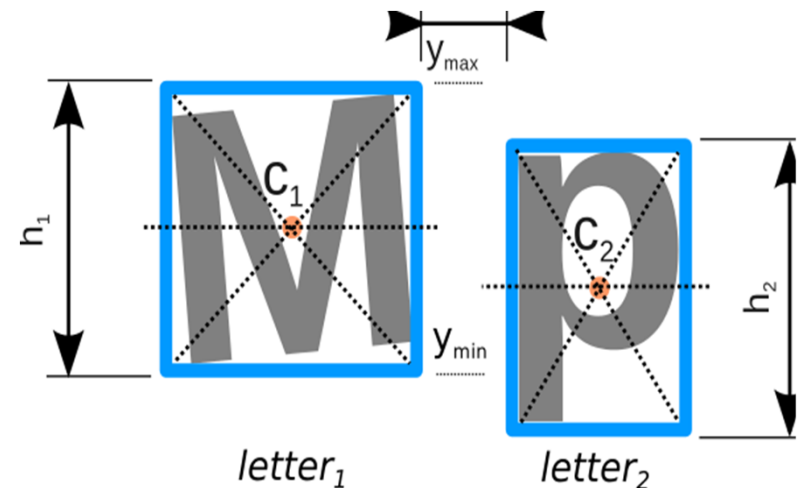
Si $\sigma > c / dm$ alors la CC est une zone candidate

Sinon la CC n'est pas une zone candidate



+ De la lettre à la ligne de texte

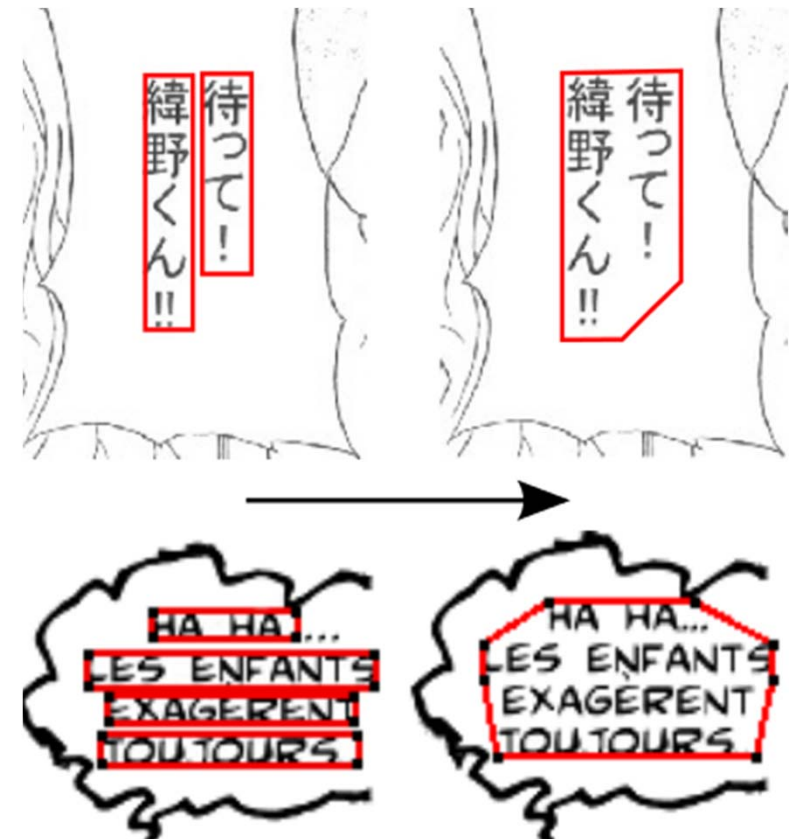
- > On considère une zone comme candidate comme la première lettre de la ligne de texte
- > On tente de regrouper les zones candidates proches si :
 - Elles ont des dimensions similaires
 - Elles sont alignées
 - L'espace entre les zones est régulier
- > Si les critères sont respectés alors les zones candidates forment une « ligne de texte »



Extraction du texte (seul)

+ De la ligne de texte au paragraphe/bloc de texte

- > Regroupement des lignes de texte
 - Hauteur (ou largeur) similaire
 - Distance interligne < hauteur d'une ligne



+ Exemples

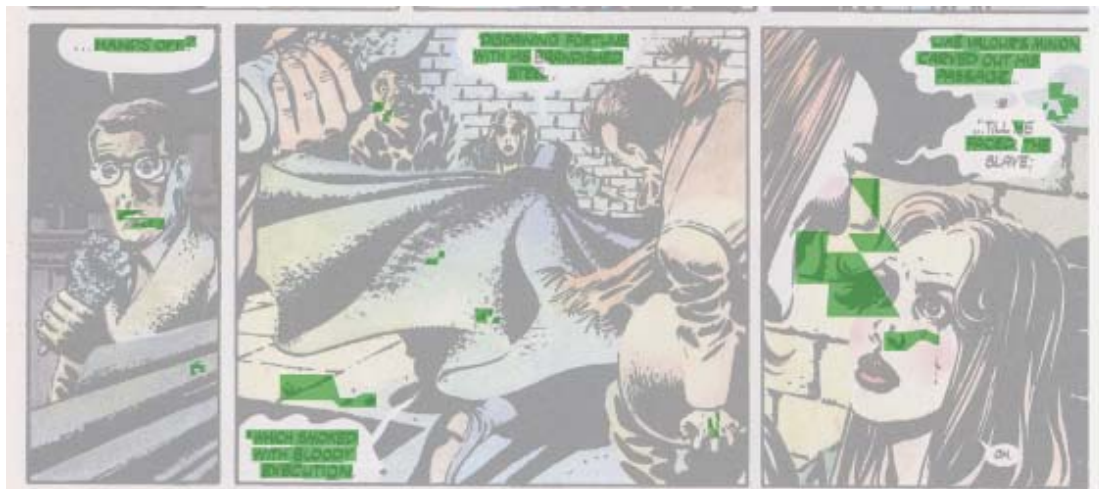


ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

Extraction du texte (seul)

+ Exemples



ED I2S : Module image 2019

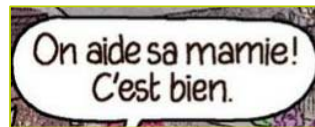
19 avril 2019 - Jean-Christophe Burie

RECONNAISSANCE DU TEXTE

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

- + Reconnaître le texte
 - > Passer de l'image à la représentation textuelle



On aide sa mamie ! C'est bien

- Indexation du texte
- Recherche plein-texte

- + MAIS : Grande diversité de polices
 - > Difficile d'envisager un OCR unique



+ Identification de 5 grandes catégories

> Groupe A

1. LA JEUNESSE
2. LA BALLADE DE LA MER SALÉE
3. LE SECRET DE TRISTAN BANTAM
4. RENDEZ-VOUS À BAHIA
5. SAMBA AVEC TIR FIXE
6. L'AIGLE DU BRÉSIL

> Groupe B

Restez en arrière et ne vous exposez pas ... Tiens ! curieux, tout paraît calme .

> Groupe C

Quand ma mère m'emmène à son travail, j'ai pas le droit de bouger parce qu'elle a peur que je fasse des bêtises dans les supermarchés .

Première approche

+ Identification de 5 grandes catégories

> Groupe D

Obsédé par la mort, celui que l'on surnommait le "Roi-Sorcier" pratiquait des rituels au cours desquels il dévorait le cœur de sacrifiés, persuadé de prolonger ainsi sa vie, et à terme, de devenir immortel.

> Groupe E

JEAN-DEXTRE MARCHE, ET PRODUIT
AINSI UNE ÉNERGIE QUI LUI SERT
À MARCHER ! IL SEMBLE QU'EXISTER
SOIT SON SEUL MÉTIER. CAR C'EST
UN TRAVAIL, L'EXISTENCE ! IL EST FACILE
NE PAS RISQUER SA VIE DE PEUR
DES MORTS-VIVANTS

+ Difficulté

Groupe A



Groupe E

-

+

+ Approches utilisées

- > Développement de prétraitement des images de texte
- > Correction des défauts pour optimiser la reconnaissance

+ Reconnaissance

- > Groupe A et B
 - Test de OCR classique, par exemple : Tesseract, Fine Reader
 - Pas d'apprentissage ou apprentissage limité
- > Groupe C
 - OCR opus
 - Apprentissage de la « police de caractère » nécessaire

+ Indice de confiance

- > Etiquetage des mots

Première approche

+ Exemples de résultats (OCR classique)

> Confiance de 90 % à 100 %

On aide sa mamie! C'est bien.



> Confiance de 50 % à 90 %

Il s'appelle AR'IOL Dis bonjour à madame PROUX



> Confiance de 0 % à 50 %

uméro Z'r'! GRiOTT E et BARTAVELLE! a —/—, \

Alors, petifzs, raconl'25 yUJWQ œque vrv v " n n: ml? |

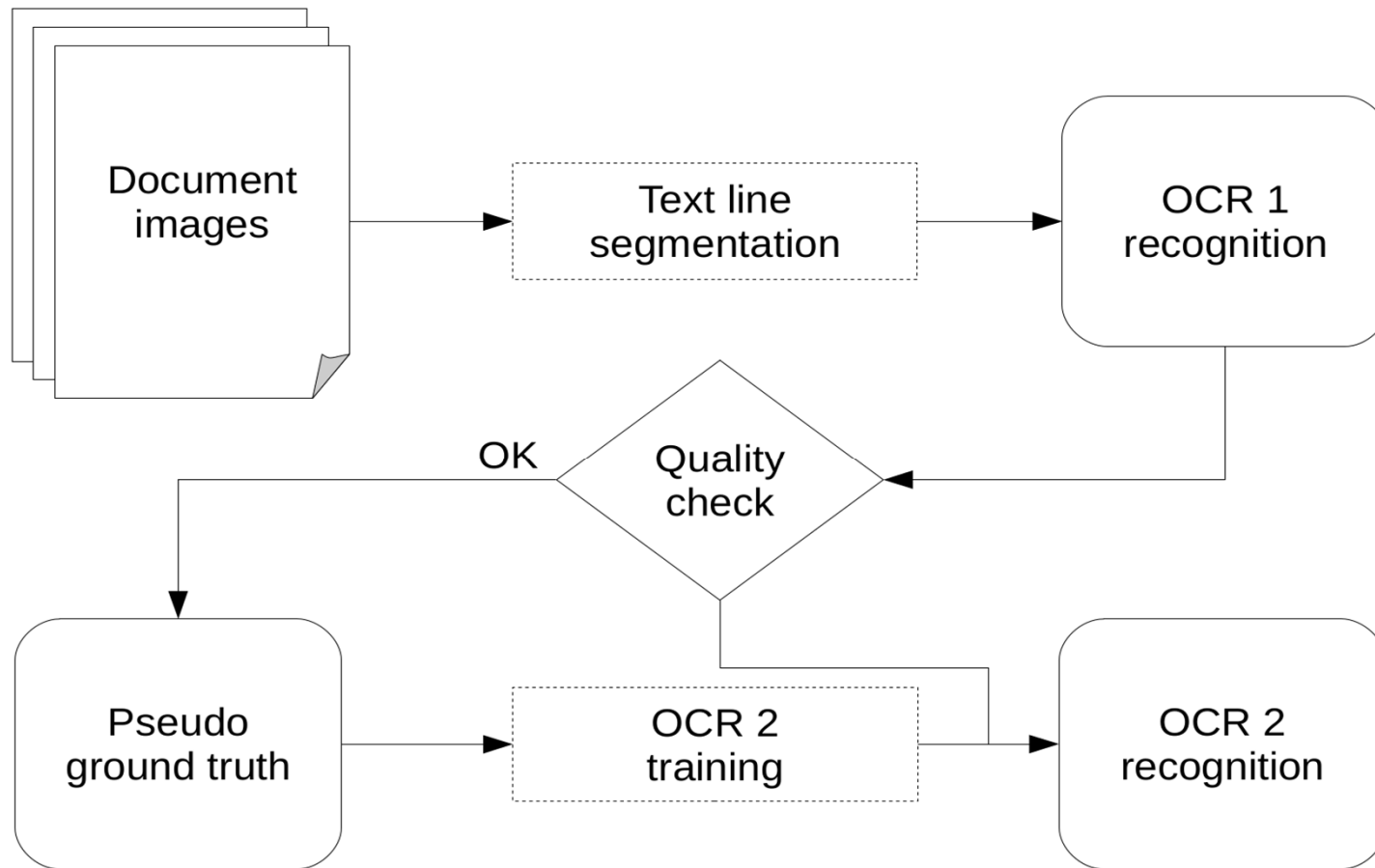


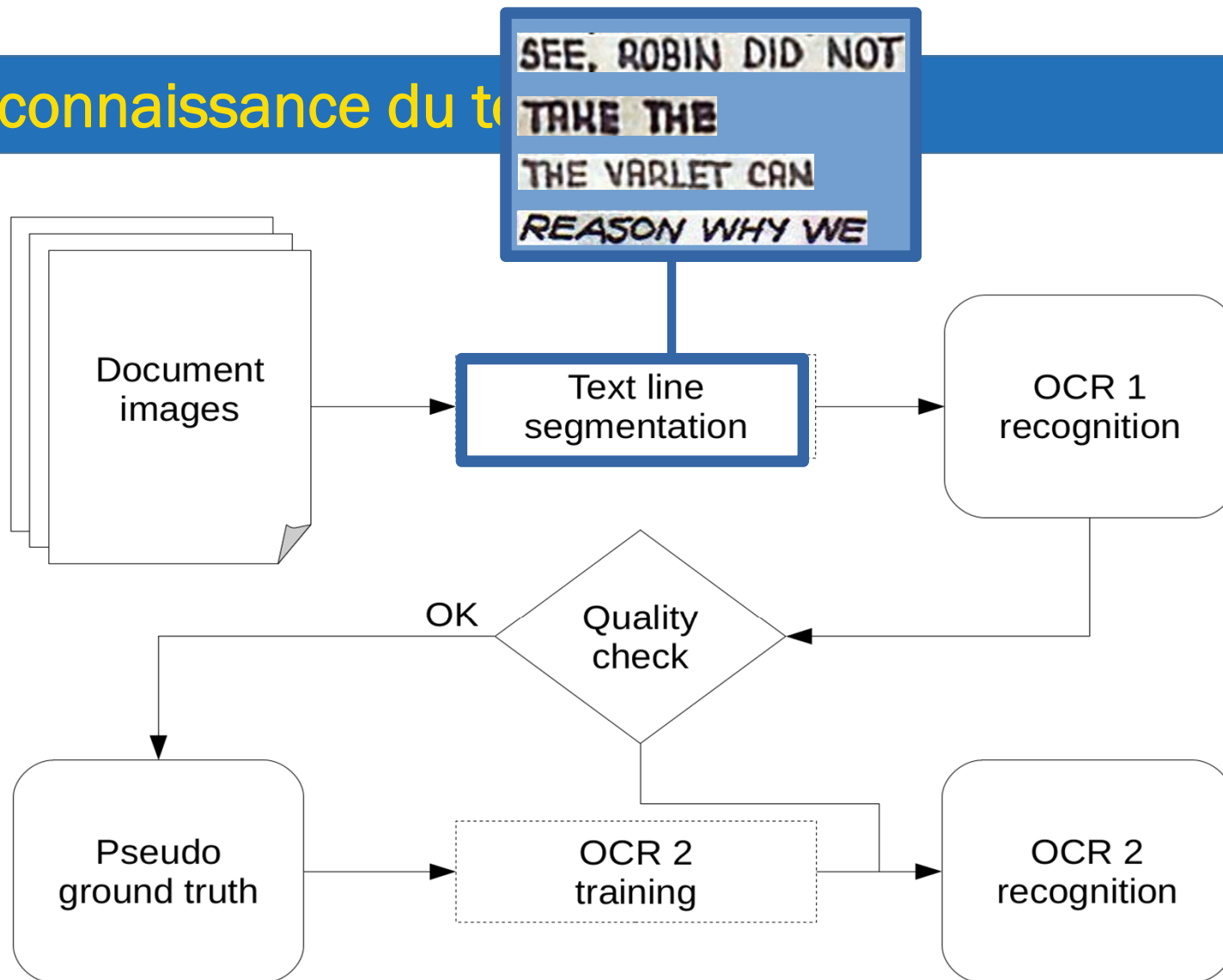
Limites des approches classiques

- + OCR « classique »
 - > Fonctionne sur un nombre de police limitée

- + OCR basés sur apprentissage
 - > Nécessite un nouvel apprentissage pour toute nouveau style
 - > Nécessite de refaire une vérité terrain
 - Processus long

Reconnaissance du texte

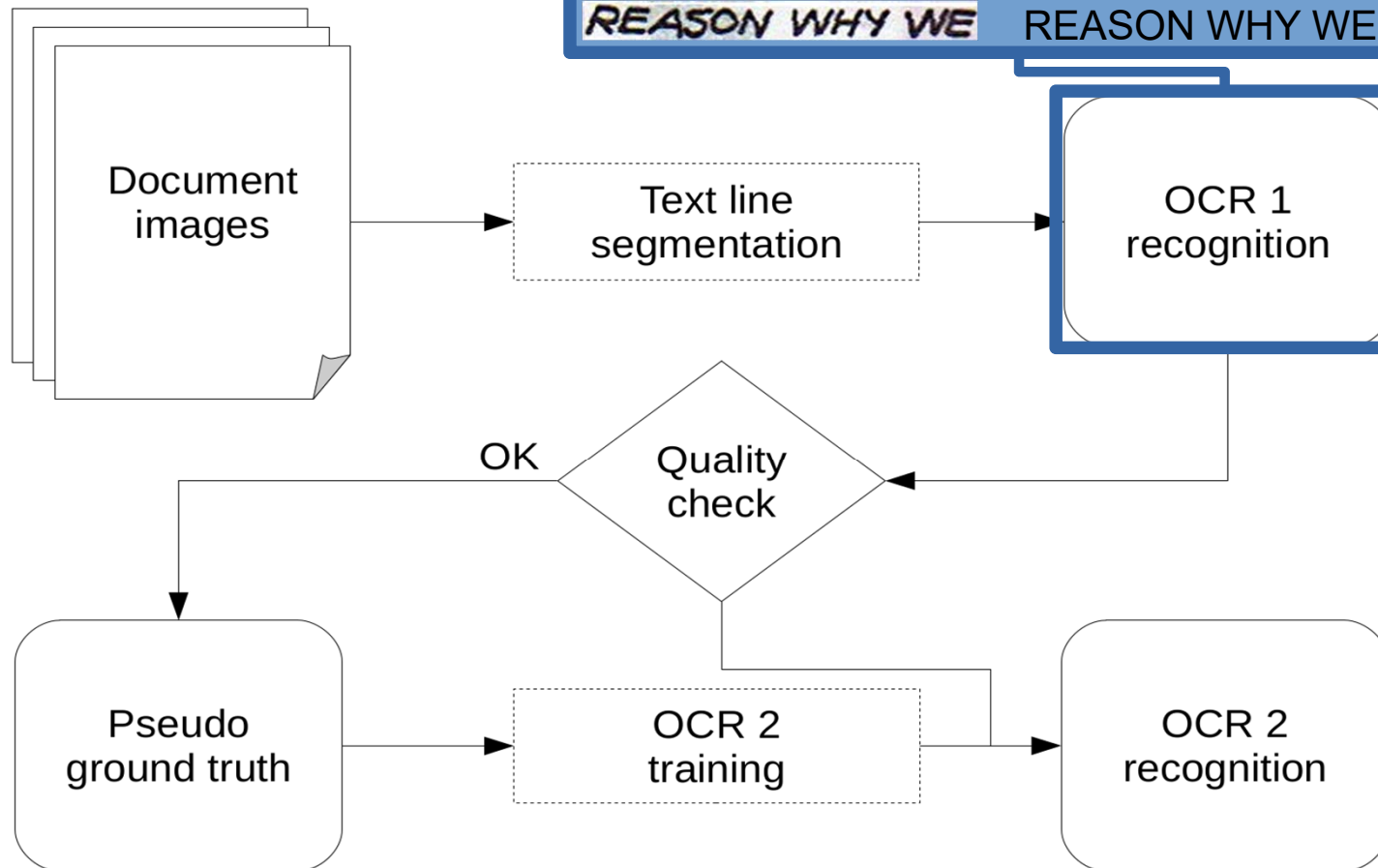




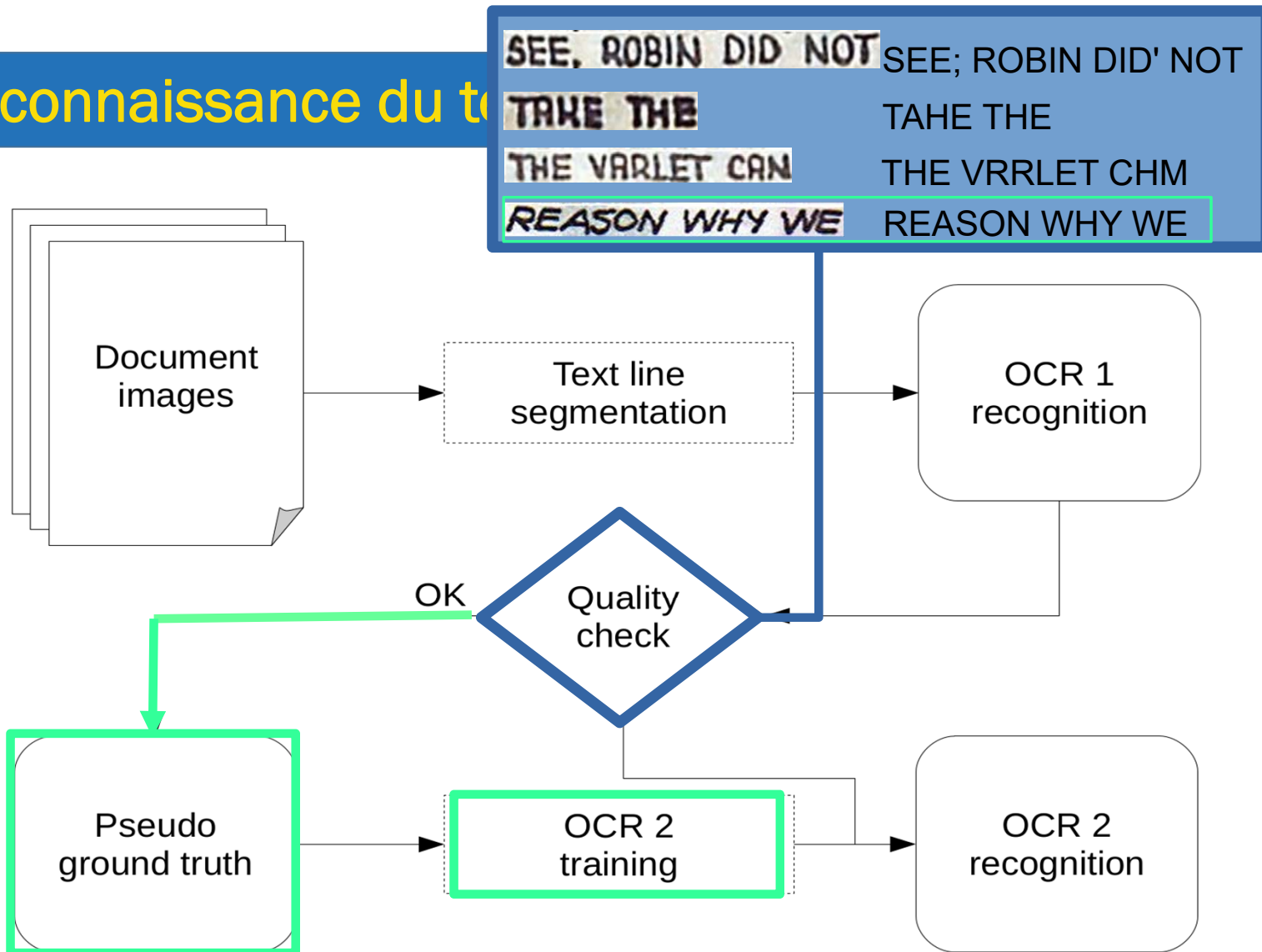
Reconnaissance du texte

SEE, ROBIN DID NOT
TAKE THE
THE VARLET CAN
REASON WHY WE

SEE; ROBIN DID' NOT
TAHE THE
THE VRRLET CHM
REASON WHY WE



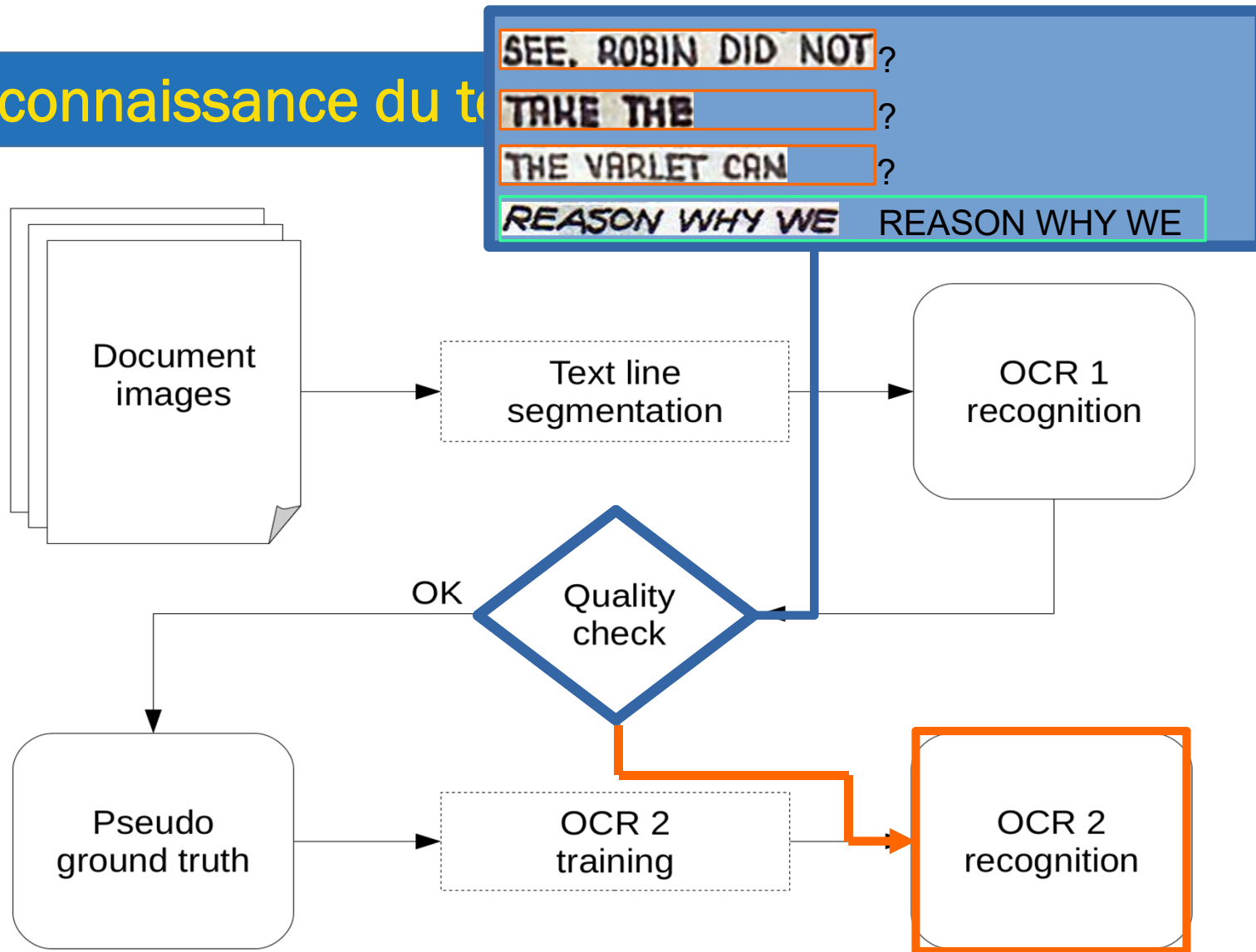
Reconnaissance du texte



ED I2S : Module image 2019

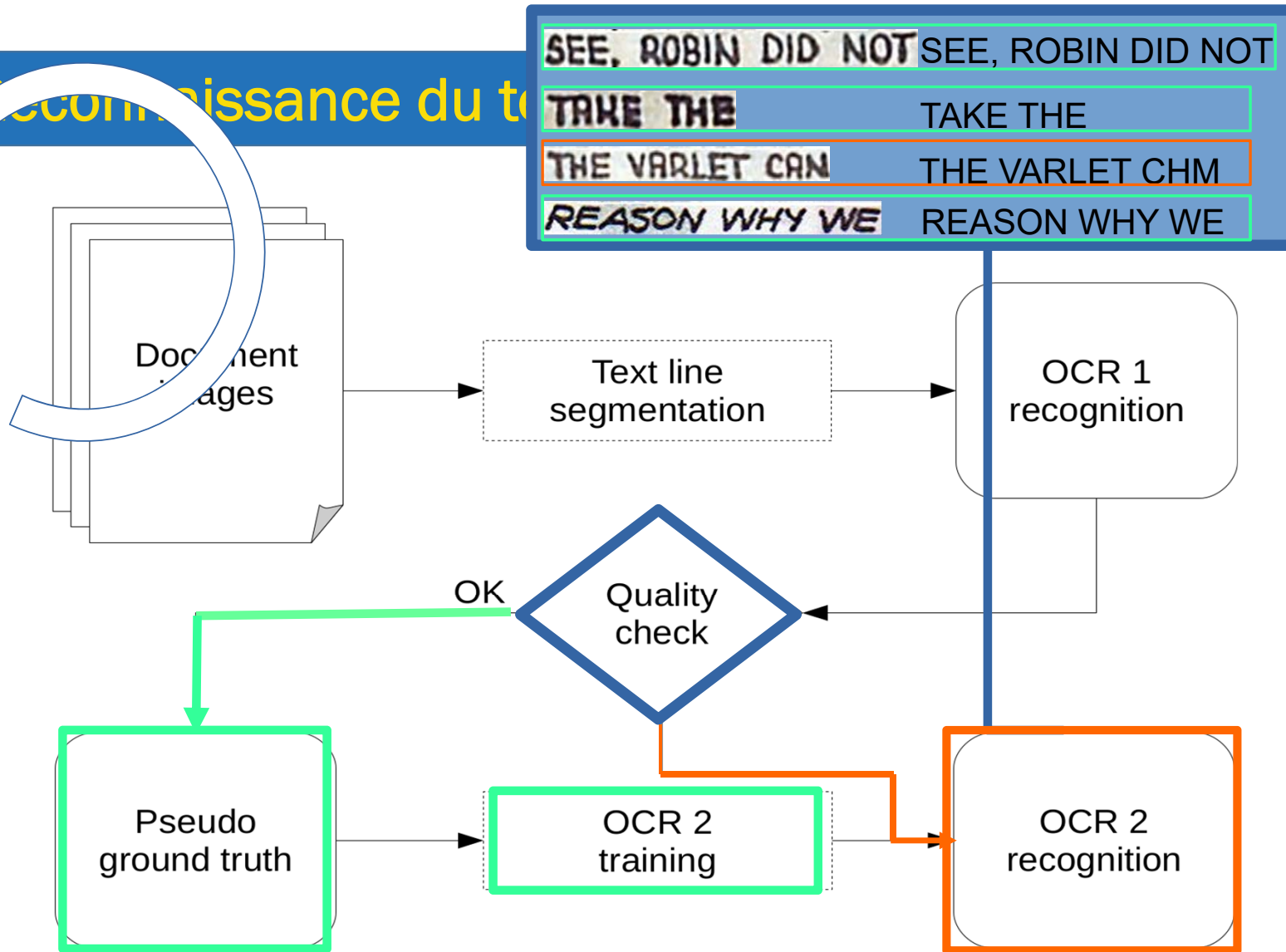
19 avril 2019 - Jean-Christophe Burie

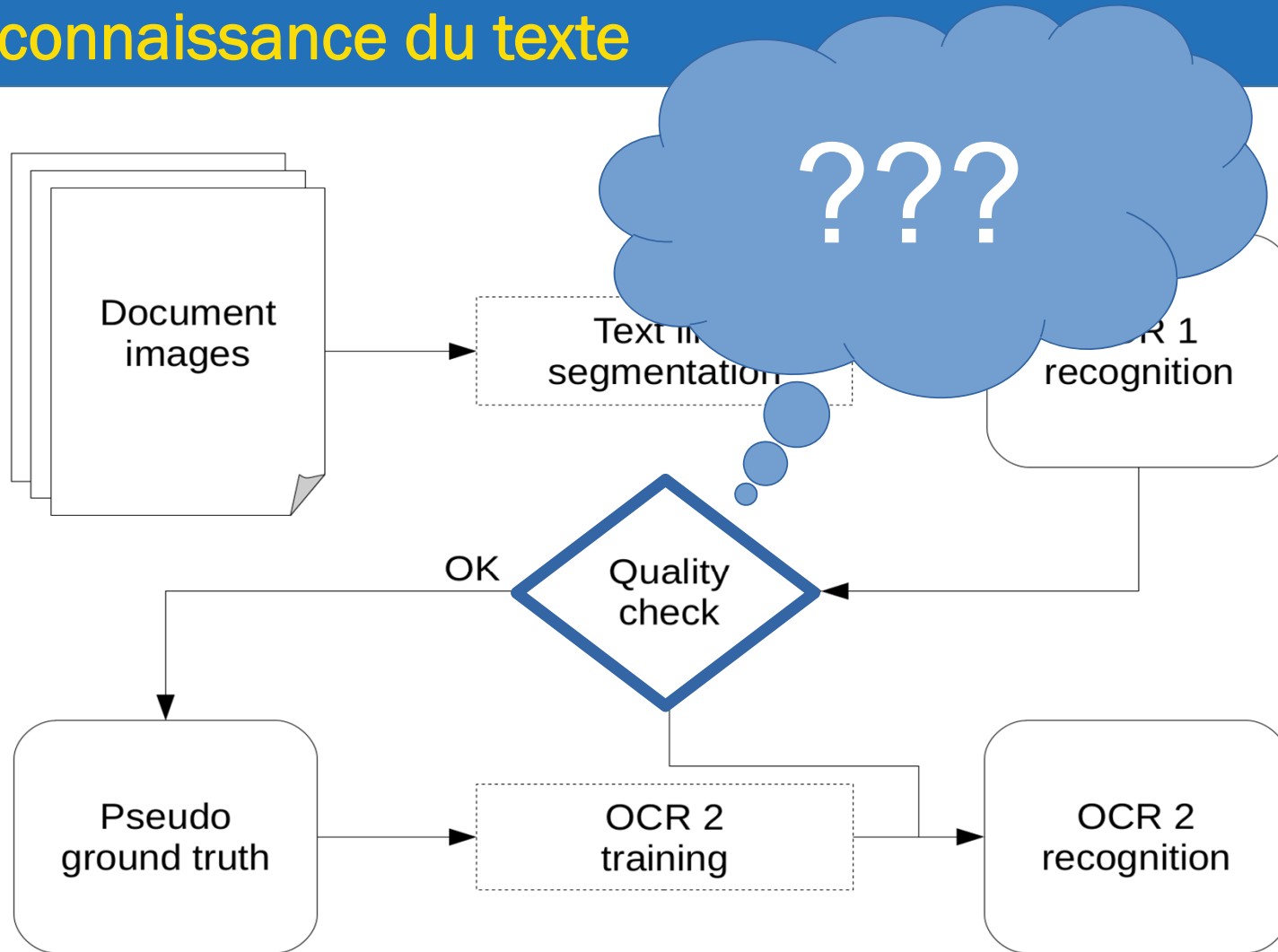
Reconnaissance du texte



ED I2

19 avril 2019 - Jean-Christophe Burie





- + Quality check = test de qualité Springmann [16]
 - > L : Lexicalité moyenne
- + $L = 1 - \text{moyenne (distance de Levenshtein par caractère)}$
- + Exemple
 - > Que vaut L pour la phrase "TAHE THE" ?
 - > 1^{ère} passe: le mot le plus proche de "TAHE" est "TAKE"
 - > 2nd passe : 7 caractères, $\text{dist}(\text{TAHE}) = 1$, $\text{dist}(\text{THE}) = 0$
 - > $L = 1 - (1+0)/7$

• L =

85.71%

[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

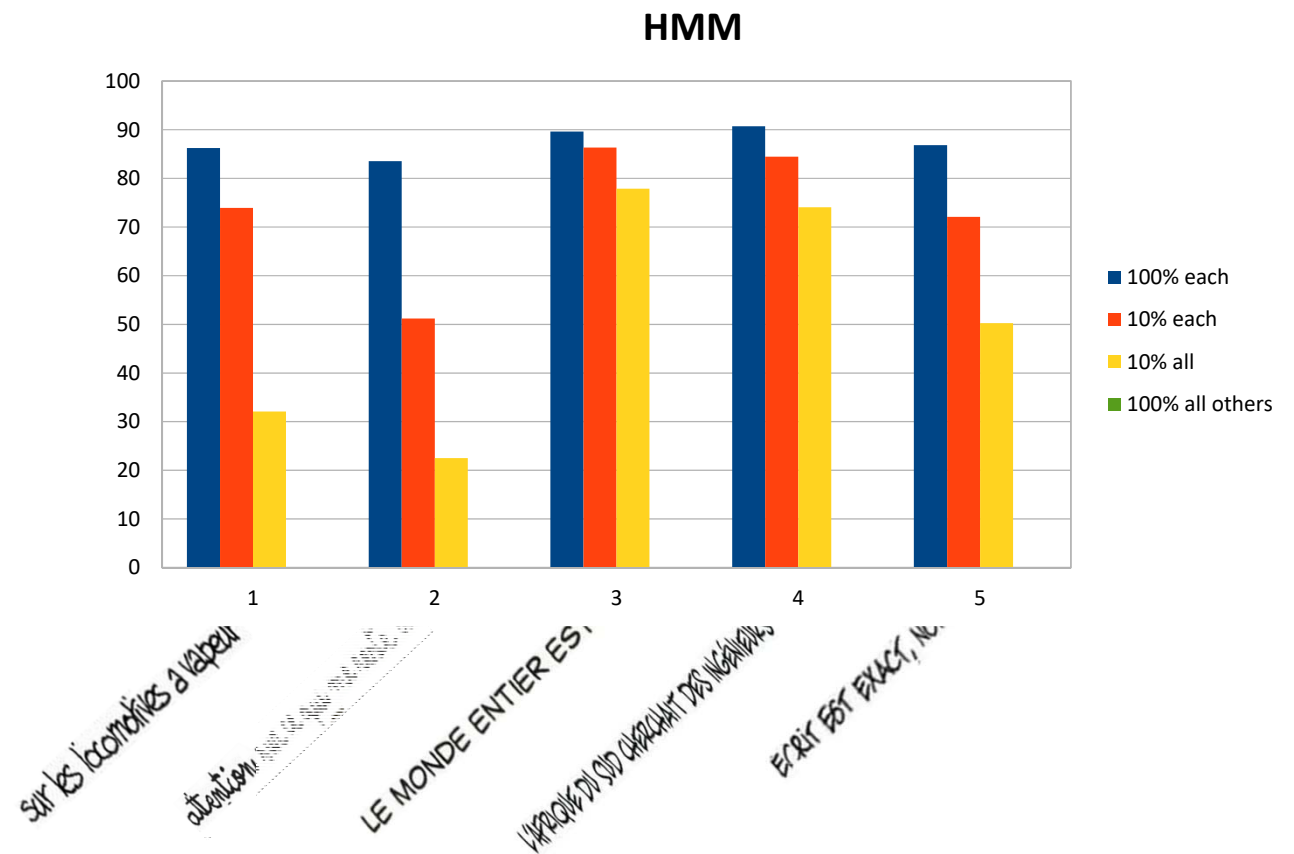
+ OCR Standard

- > Utilisation du logiciel libre Tesseract
 - Capable de reconnaître ~60 styles d'écritures dactylographiés
 - Dans 101 langues.
- > Performance

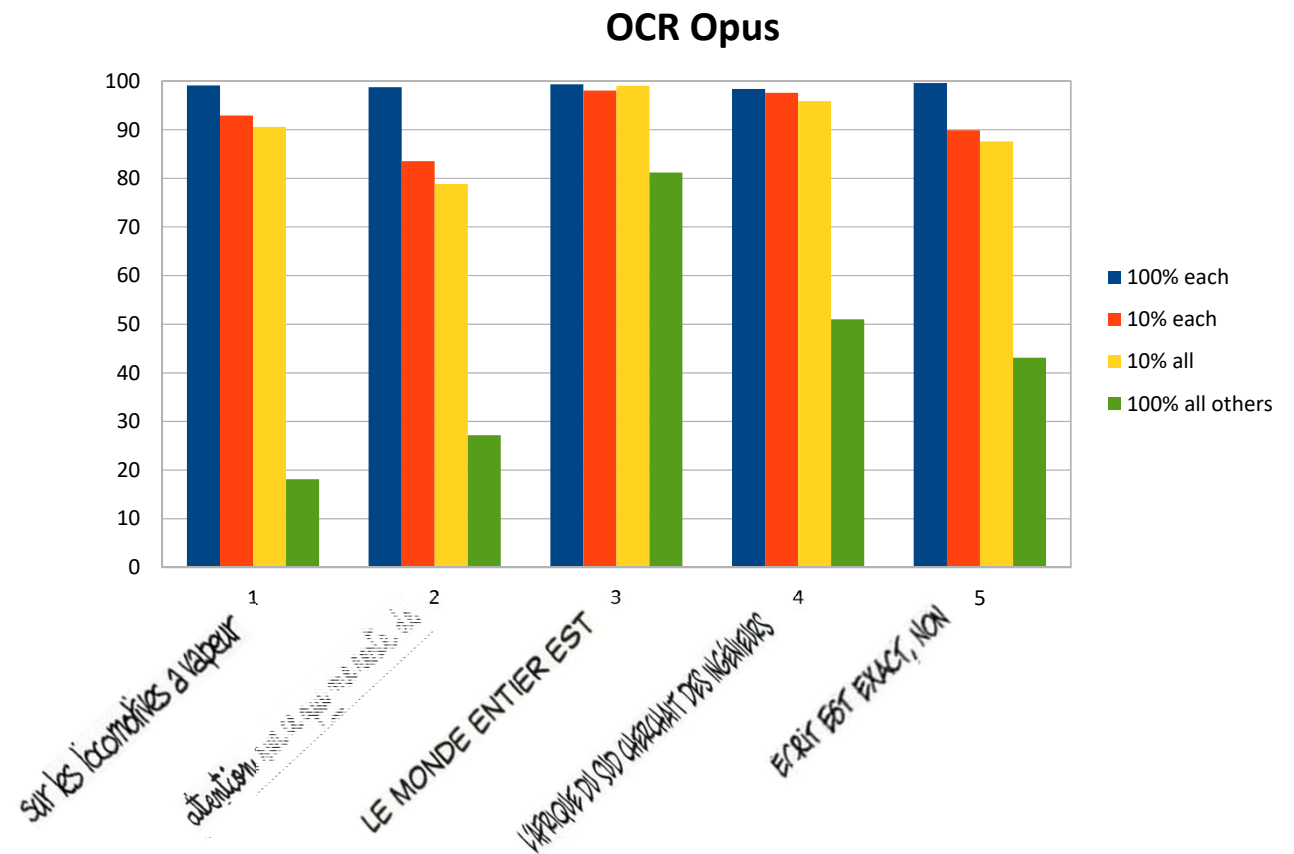
Album	Taux de reconnaissance (%)
9782203073760 <i>sur les locomotives à vapeur</i>	47.27
9782754808323 <i>attention sur ce qui me reste de</i>	19.12
9782369812500 <i>LE MONDE ENTIER EST</i>	86.80
9782754815925 <i>L'AFRIQUE DU SUD CHERCHAIT DES INGÉNIEURS</i>	40.53
9782800176703 <i>ECRIT EST EXACT, NON</i>	19.44

- + OCR + Apprentissage manuel
 - > Comparaison de deux algorithmes
 - HMM (modèle de Markov caché)
 - OCRopus (réseaux de neurones récurrents)
- + Principes de fonctionnement
 - > Préparation des données d'apprentissage
 - Sélection d'images de lignes de texte et transcriptions manuelles (5x500 images)
- + Apprentissage par l'algorithme
 - > Test des performances suivant trois scénarios d'apprentissage
 - 100 % du même album
 - 10 % du même album
 - 10 % de tous les albums
 - 100 % de tous les autres albums (OCRopus seulement)

- + OCR + Apprentissage manuel
- > Résultat



+ OCR + Apprentissage manuel > Résultat



+ OCR + Apprentissage automatique

- > Utilisation d'OCRopus
 - entraîné sur 500 lignes de texte de 5 albums + eBDtheque
 - 6189 ligne au total
- > Apprentissage des nouvelles lignes reconnues après chaque itération
 - ligne reconnue = tous ses mots présents dans le lexique de la langue

+ OCR + Apprentissage automatique

> Résultats

Itération 1



Itération 3

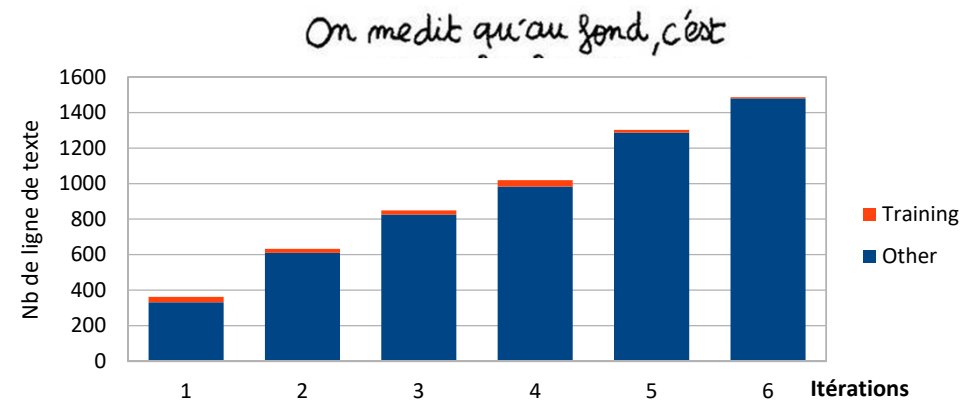
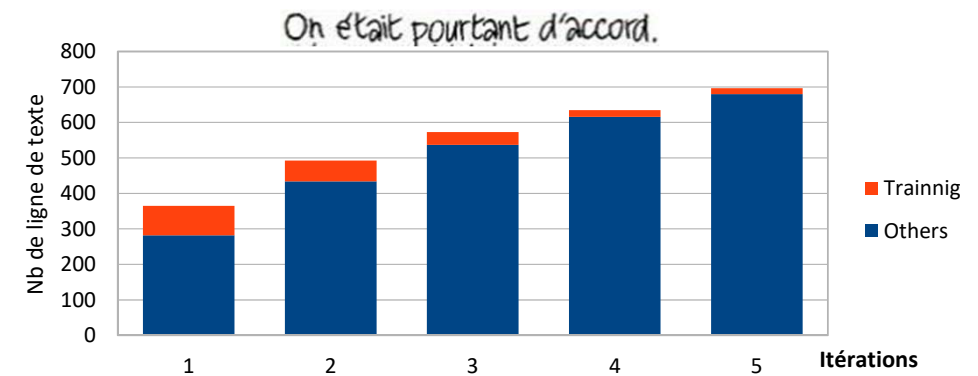


ED I2S : Module image 2

19 avril 2019 - Jean-Christophe Burie

+ OCR + Apprentissage automatique > Résultats

Ligne de texte ok =
mots faisant partie
du lexique



+ Conclusions

- > OCR sans apprentissage : pour texte dactylographié
- > OCR avec apprentissage manuel : pour texte manuscrit
- > OCR avec apprentissage automatique :
 - pour certain styles d'écriture manuscrite

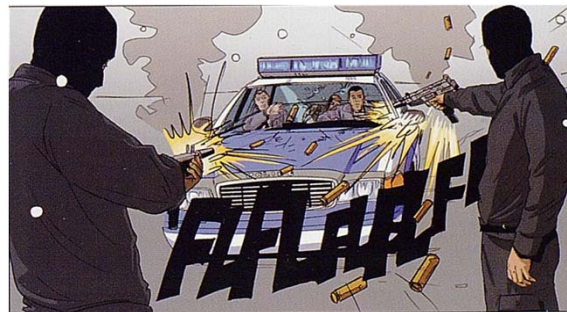
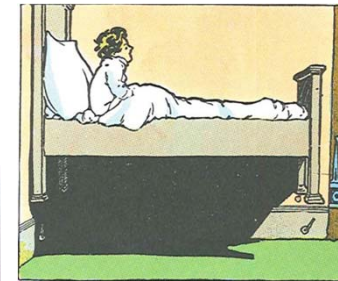
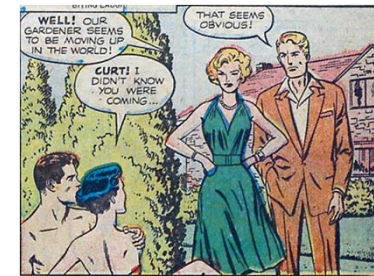
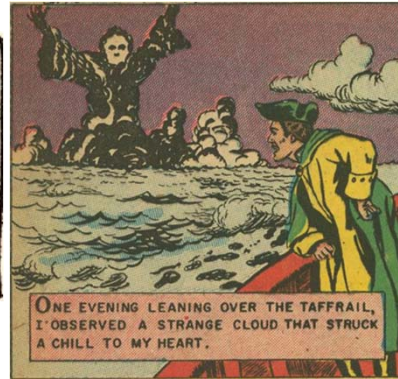
+ Il reste encore beaucoup à faire ...

DÉTECTION DES PERSONNAGES

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

- + Grande diversité dans la représentation des personnages



ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

APPROCHE STATISTIQUE « COULEUR »

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ CBIR : Content Based Image Retrieval

- > Recherche d'images par le contenu
- > Comment rechercher les personnages ?
- > Quel(s) critère(s) ?

→ couleur semble être une constante

+ Principe de la méthode :

- > Réduction du nombre de couleurs
- > Obtenir les couleurs de l'image requête (interaction avec l'utilisateur)
- > Calculer le descripteur correspondant à la requête de l'utilisateur
- > Rechercher les instances de personnages similaires



+ Réduction des couleurs



N > 10000

N = 256

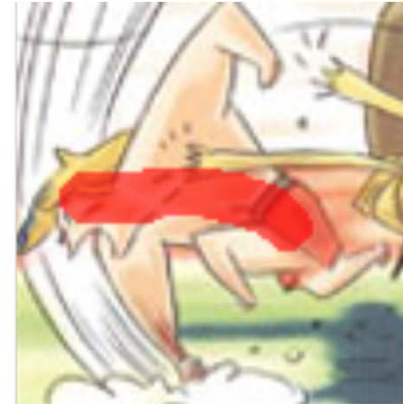
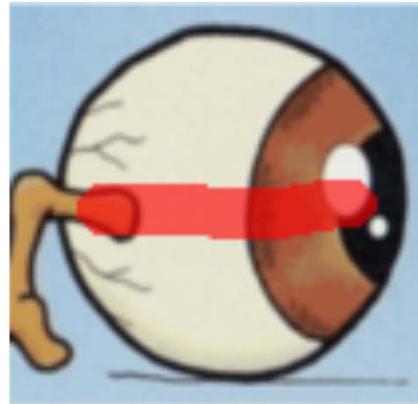
N = 32

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

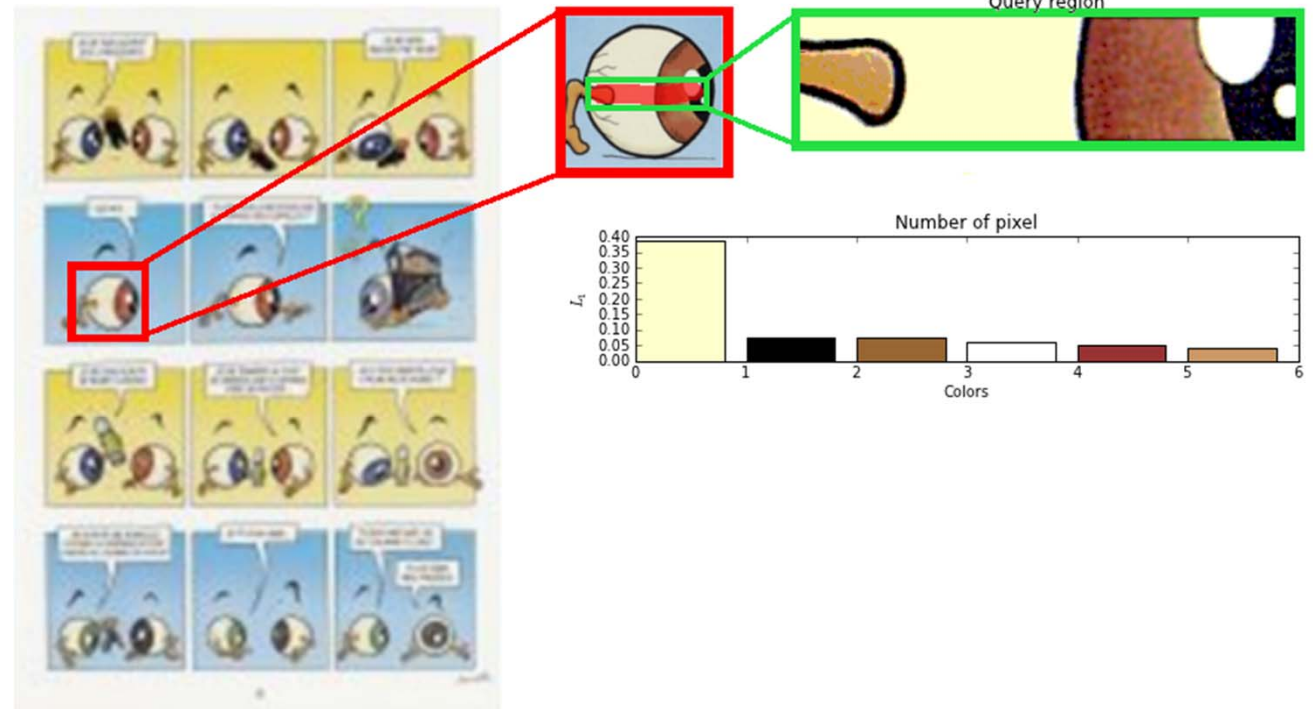
Approche statistique "couleur"

+ Requête utilisateur



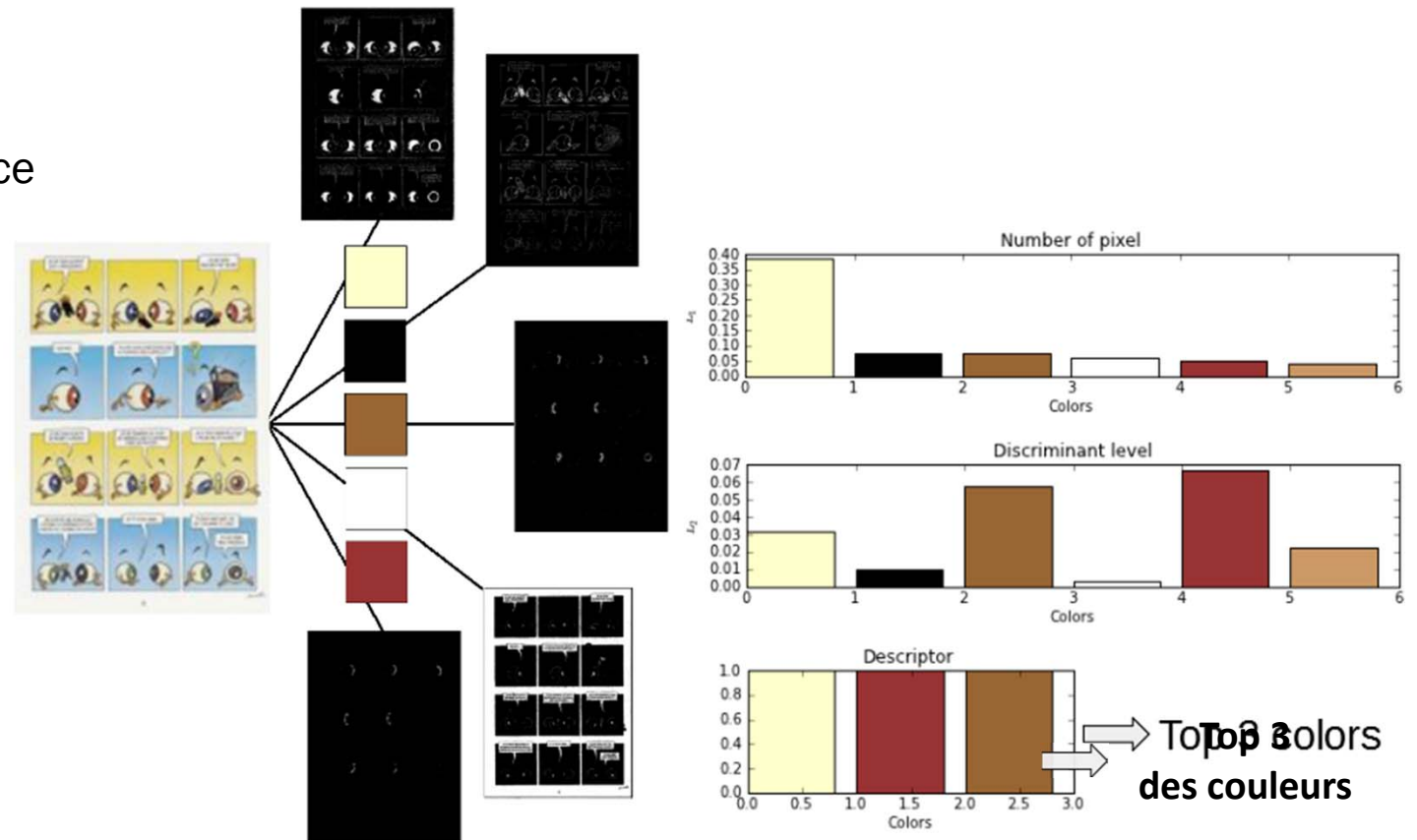
Approche statistique "couleur"

- + Sélection des couleurs
- Nombre
d'occurrence de
chaque couleur



Approche statistique "couleur"

- + Analyse de couleur :
 - > Occurrence
 - > Niveau de discriminance



Approche statistique "couleur"

+ Exemple



Entrée utilisateur



Descripteur

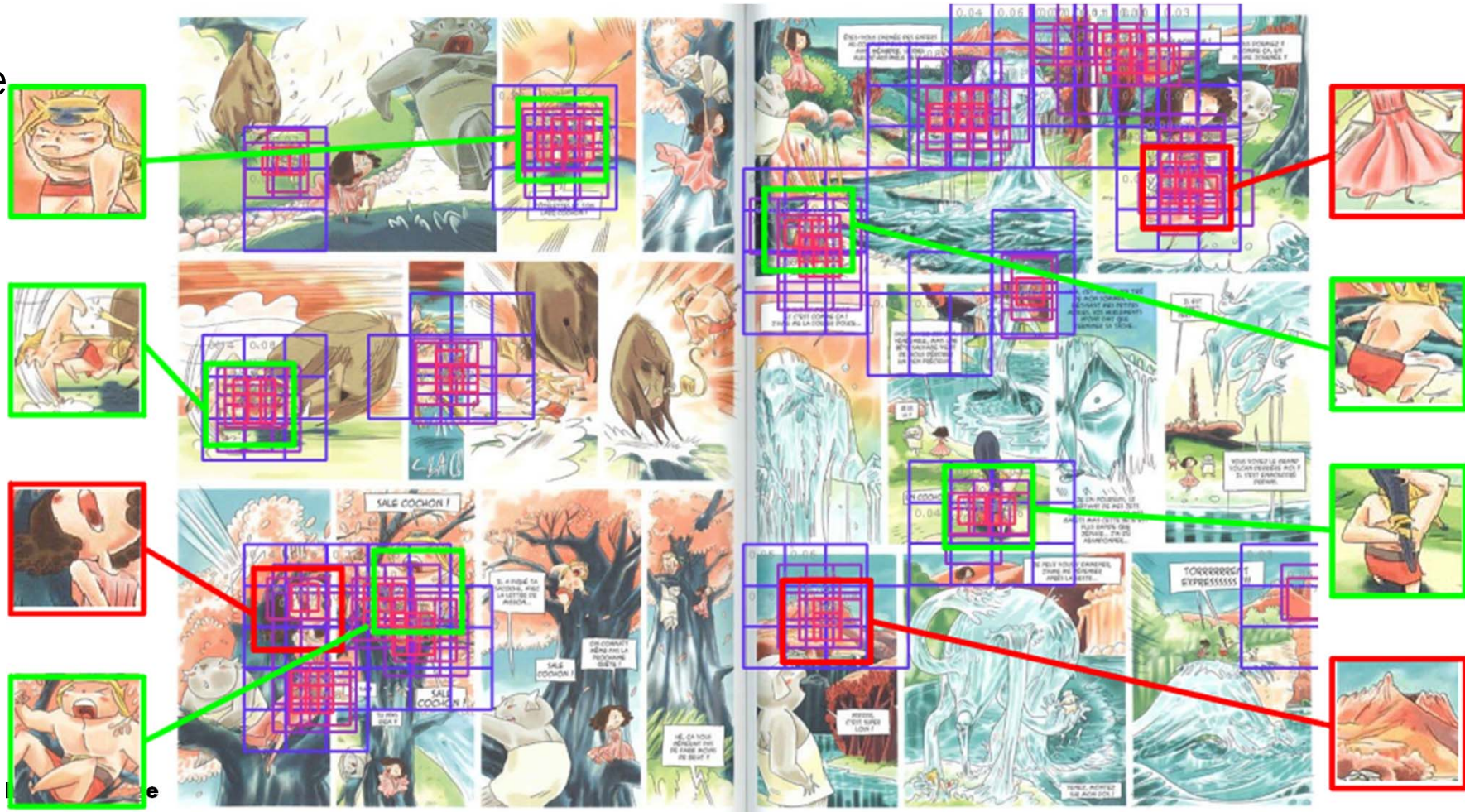


ED I2S : Module image 2

19 avril 2019 - Jean-Christophe Burie

Approche statistique "couleur"

+ Exemple



ED I2S : I2S e

19 avril 2019 - Jean-Christophe Burie

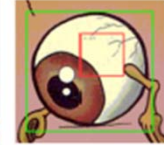
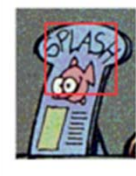
Approche statistique "couleur"

+ Conclusion

- > Content Based Drawing Retrieval
- > Méthode robuste au changement d'échelle, à la rotation, la translation, la déformation, les occlusions
- > Méthode profite de la redondance d'information

+ Limitations

- > Seulement pour les BD en couleur
- > Réduction du nombre de couleur (quantification)
- > Nécessite un exemple de chaque caractère
- > Perte des relations spatiales

Query	Correct	Wrong
		
		
		
		
		
		

FOUILLE ET MISE EN CORRESPONDANCE DE GRAPHES

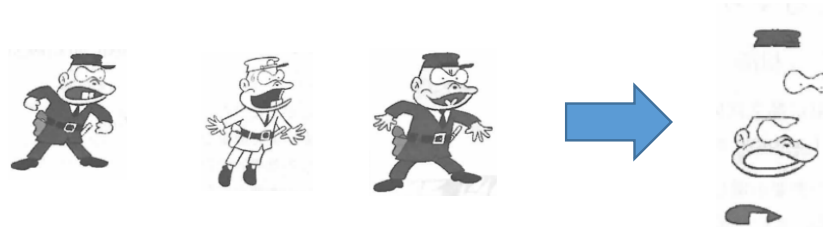
ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Thèse de Nam Thanh LE

+ Principe

> Utiliser des caractéristiques distinctives répétitives pour la « reconnaissance »



Certains éléments distinctifs sont facilement reconnu sur le visage du personnage

> Représenter les BD par des graphes attribués

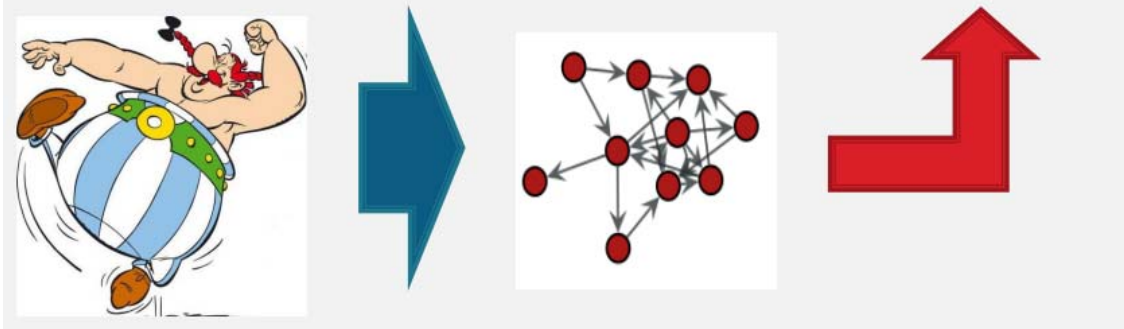
- Encoder la structure, la topologie, la géométrie, les propriétés symbolique/numérique
- Ne pas perdre la structure des personnages
- Base mathématique forte et mature

+ Représentation schématique

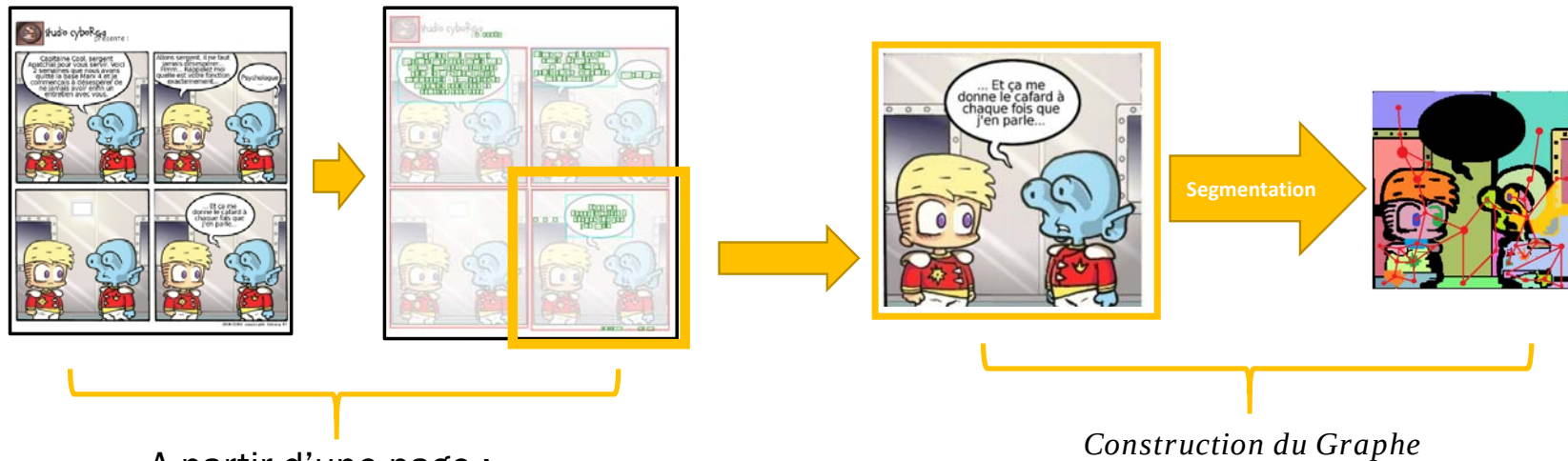
Case d'une
BD



requête



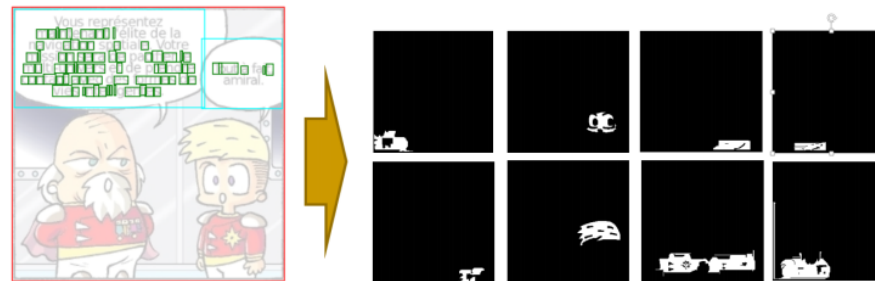
+ Construction des graphes à partir d'une page de BD



A partir d'une page :

- Pré-traitement :
 - Extraire chaque case d'une planche de BD
 - Localiser les bulles et le texte
 - Reduire le nombre de couleur
- Construction du graphe :
 - Un graphe par case

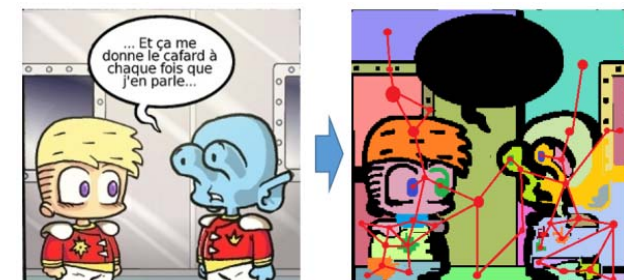
+ Construction des nœuds



+ Les nœuds correspondent aux régions segmentées

+ Information encodée dans chaque nœud

- > Couleur
- > Forme



Construction du graphe

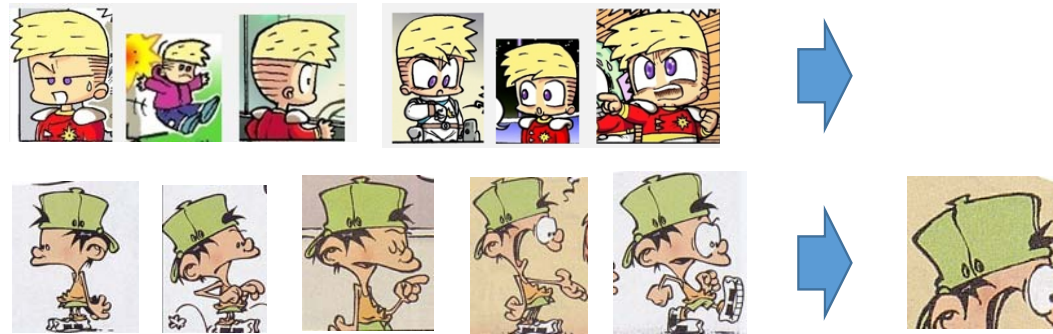
- + Information encodée dans chaque nœud
 - > Couleur



Construction du graphe

+ Information encodée dans chaque nœud

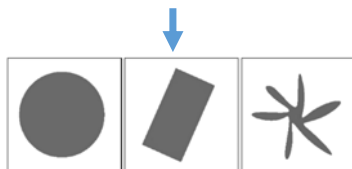
> Forme



> Descripteur de forme ?

- *Il existe différentes façons de décrire une forme*

- *Moment invariant*
- *Compacité des regions*



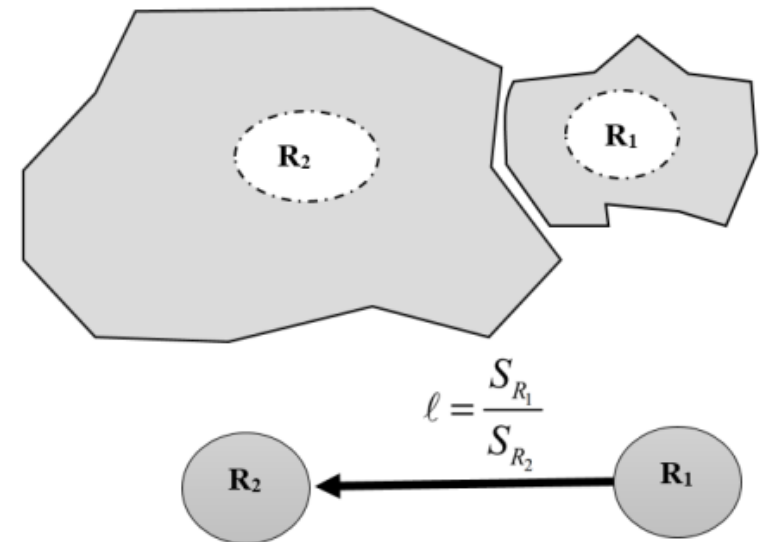
ED I2:



Shape	M1	M2	M3
S1	0.302983	0.033314	0.008469
S2	0.301319	0.032219	0.008898
S3	0.298361	0.030637	0.011182
S4	0.300351	0.032081	0.008779
S5	0.299193	0.030754	0.016486
S6	0.300596	0.032194	0.008638
S7	0.302571	0.032230	0.015247
S8	0.308926	0.035542	0.015569
S9	0.295788	0.029478	0.007937
S10	0.300463	0.031691	0.017546
S11	0.288031	0.028482	0.006141

+ Construction des arcs

- > Un arc relie deux régions R1 et R2 qui se touchent
 - graphe d'adjacence
- > L'attribut associé à l'arc représente le rapport des surfaces de chaque région

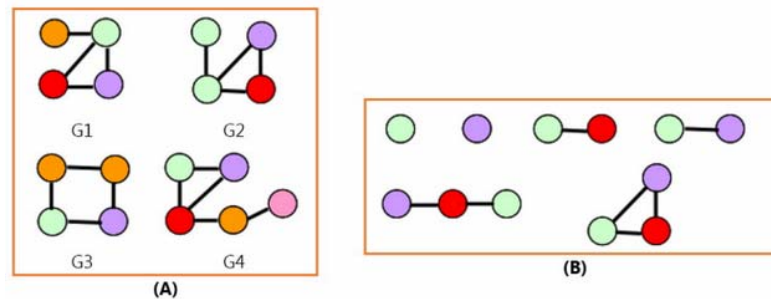


+ Problématique

- > Les graphes sont rarement identiques
- > Pas d'isomorphisme entre les graphes d'un même personnage

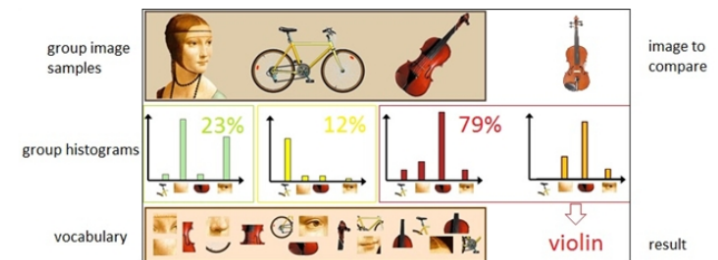
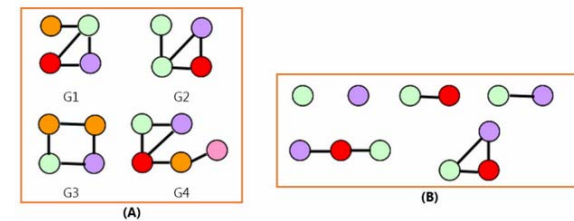
+ Principe

- > Rechercher des sous graphes fréquents



+ Objectif

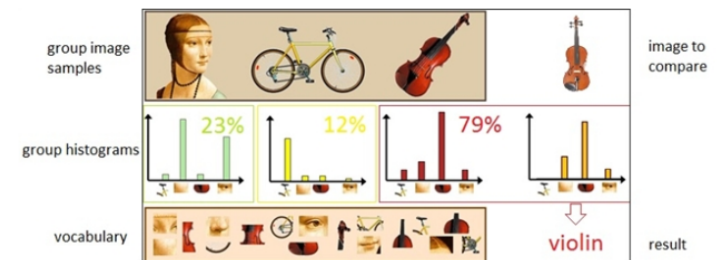
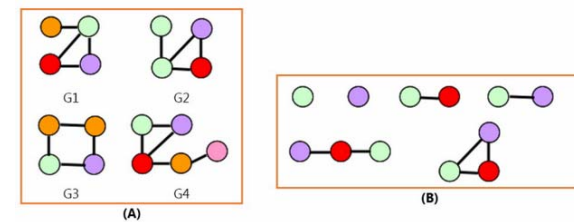
- > Fouiller les structures topologiques embarquées dans les graphes
- > Les personnages “principaux” apparaissent de manière répétée dans un album



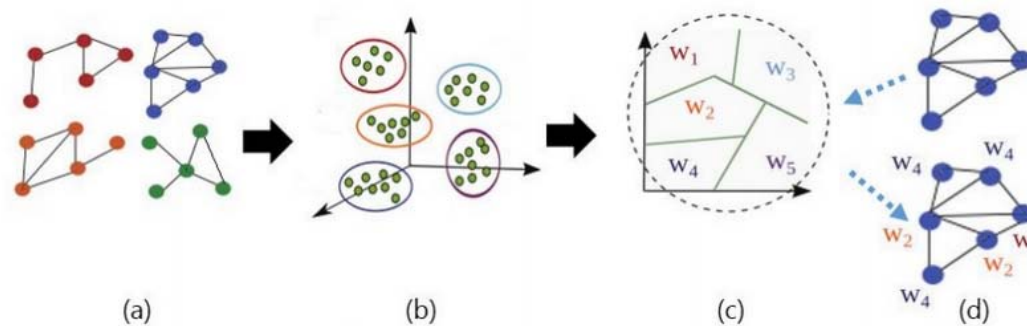
Fouille de sous graphes fréquents

+ Objectif

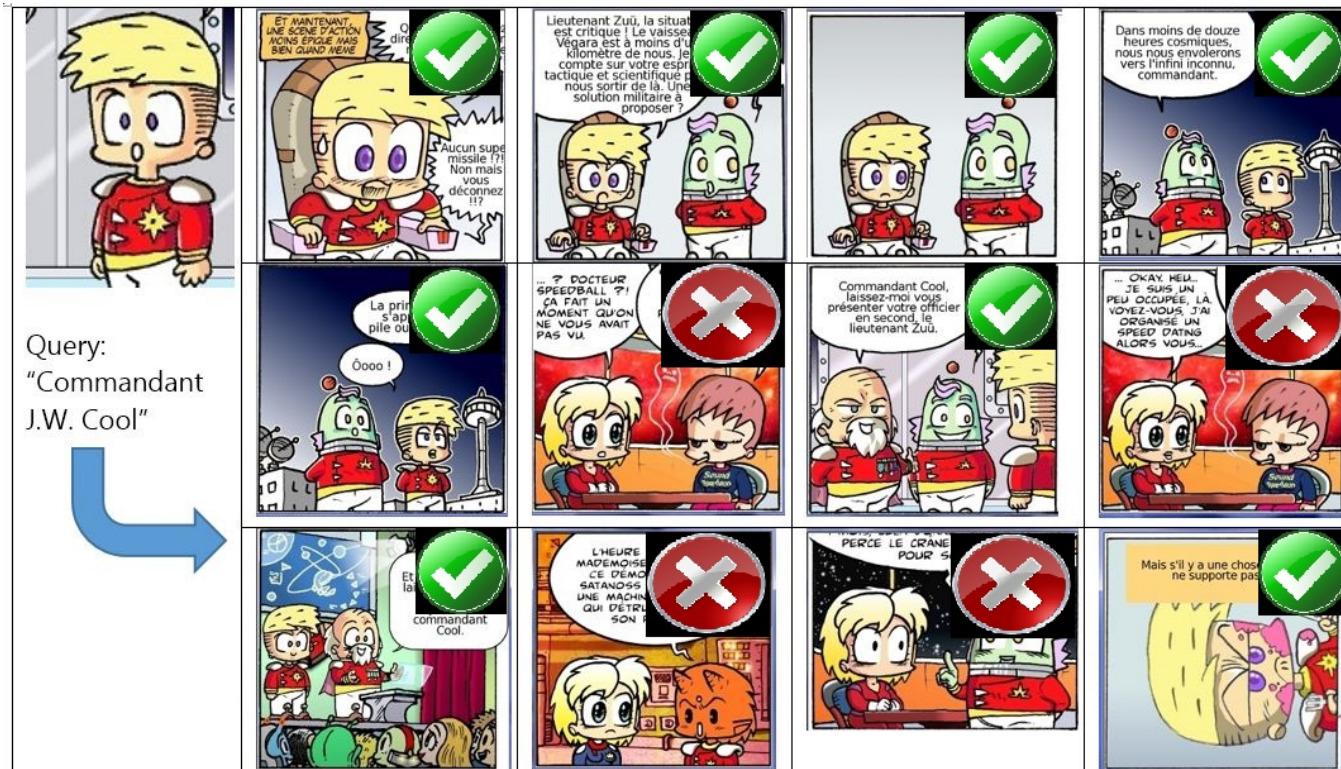
- > Fouiller les structures topologiques embarquées dans les graphes
- > Les personnages “principaux” apparaissent de manière répéter dans un album
- > Retrouver les sous-graphes avec les mêmes « propriétés »



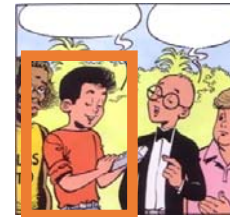
- + Analyse à différents niveaux
 - > Création de couches de sous-graphes fréquents de propriétés similaires
- + Avantage
 - > Permet de résoudre le problème d'explosion combinatoire



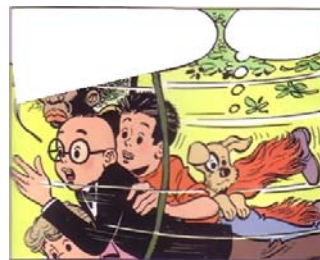
+ Exemple 1



- + Exemple 2
 - > 45 instances à trouver



- + Méthode proposée renvoie 12 résultat correcte sur 18 réponses
➔ 66.66% de bon résultat parmi les valeurs renvoyées



APPRENTISSAGES PROFONDS

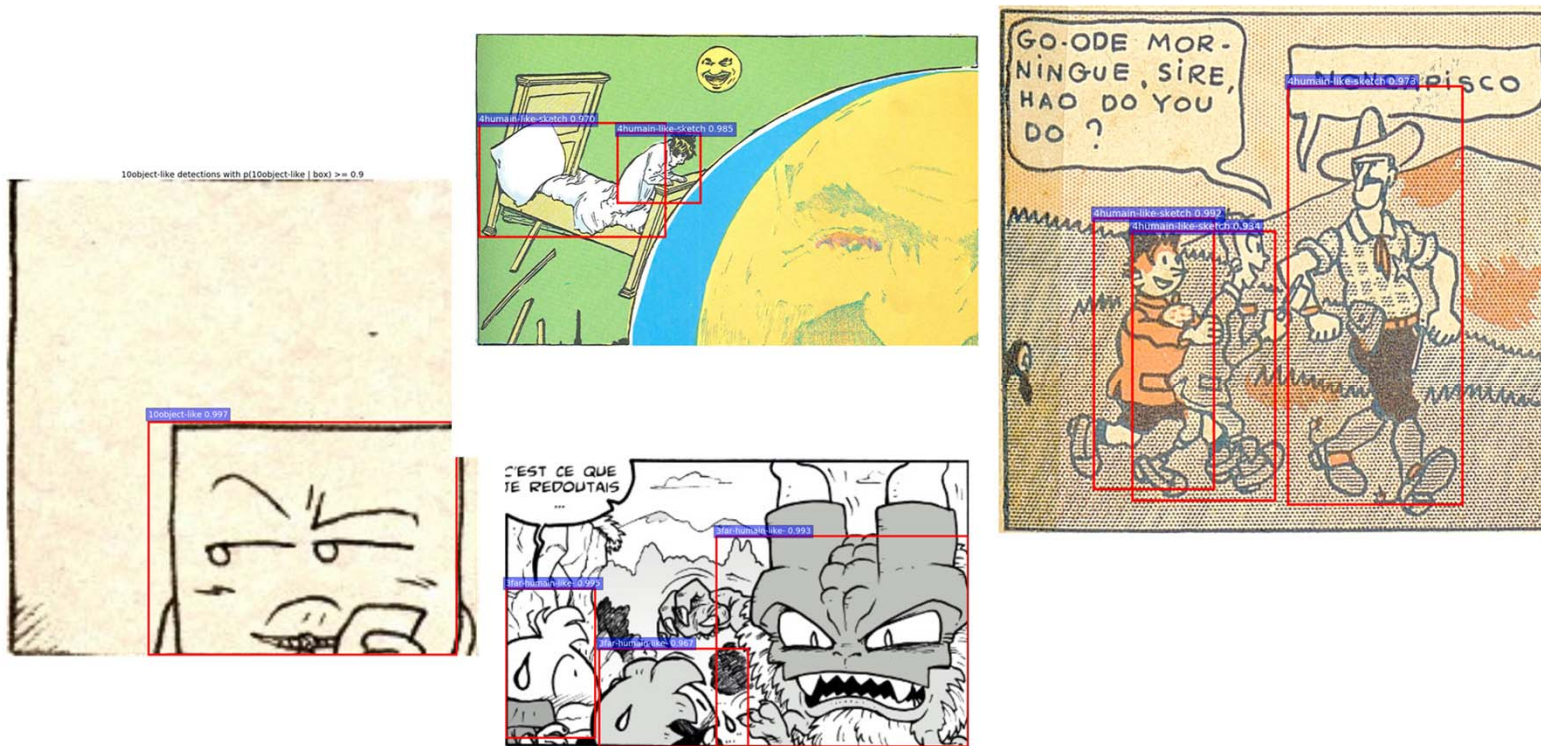
ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

Etape 1 : Détection de personnages

- + Etape 1 : Deep learning pour la détection de personnages de BD
 - > Stratégie
 - Phase 1 : Détection des personnages
 - Phase 2 : Détection des visages
 - > Test de différentes configurations
 - Faster-RCNN
 - YOLO
 - > Besoin de données pour l'apprentissage
 - Annotation manuelle
 - Catégories : "1human", "2human-like", "3near-human-like", "4far-human-like", ...
 - Jeux de données : eBDthèque + PascalVOC

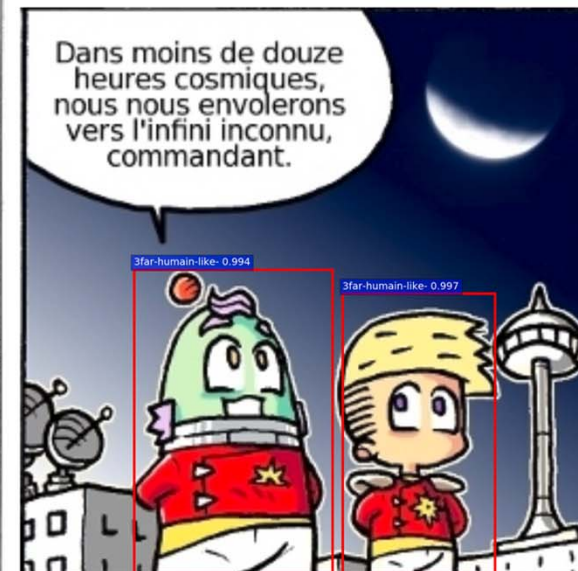
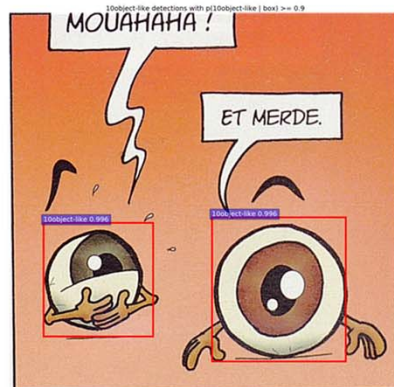
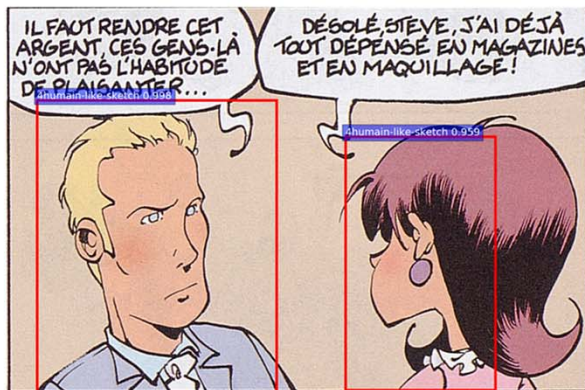
Etape 1 : Détection de personnages



ED I2S : Module im

19 avril 2019 - Jean-Christophe Burie

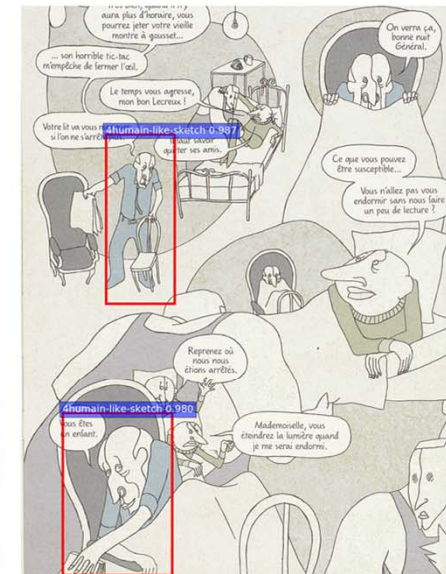
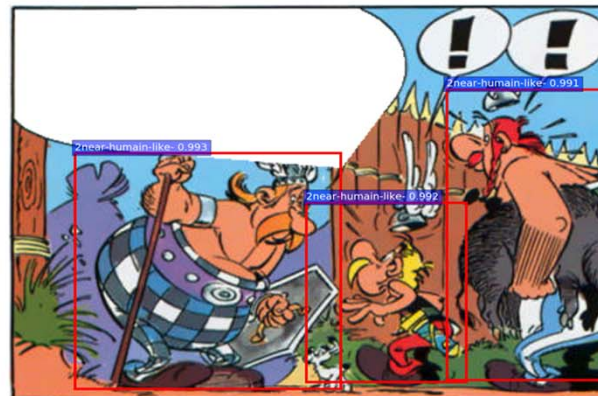
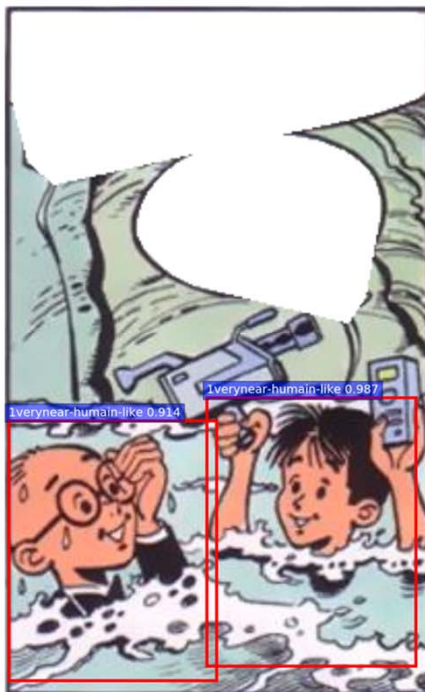
Etape 1 : Détection de personnages



ED 12S : Module Image 2019

19 avril 2019 - Jean-Christophe Burie

Etape 1 : Détection de personnages



ED I2S : Module image 2019

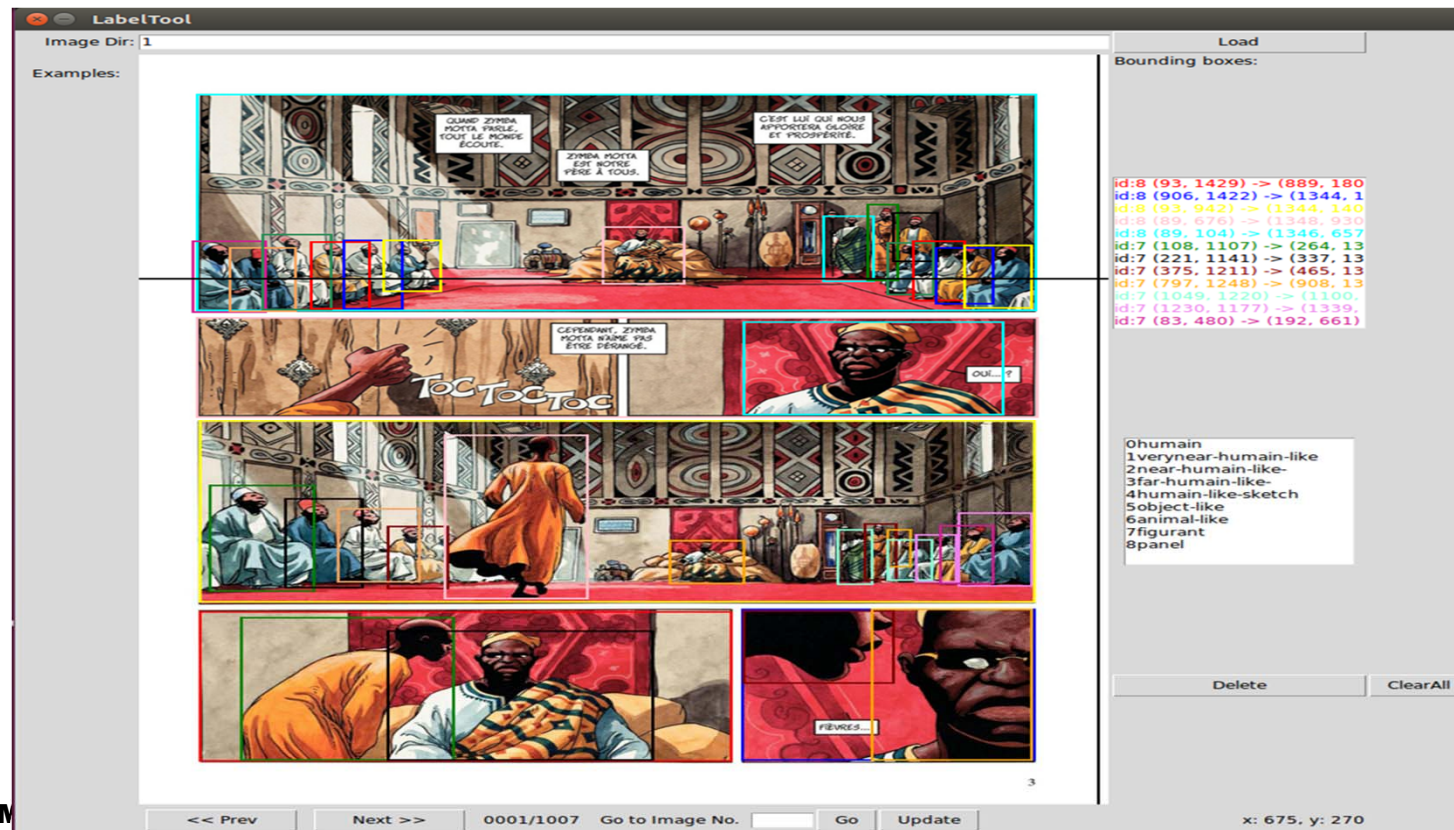
19 avril 2019 - Jean-Christophe Burie

Etape 2 : Détection de personnages

- + Etape 2 : Deep learning pour la détection de personnages de BD
 - > Test de différentes configurations
 - Faster-RCNN
 - YOLO
 - > Nouveau jeu de données ➔ Images de Sequencity
 - Choix de 10 000 pages
 - Annotation manuelle de 612 pages

Etape 2 : Détection de personnages

+ Nouveau jeu de données

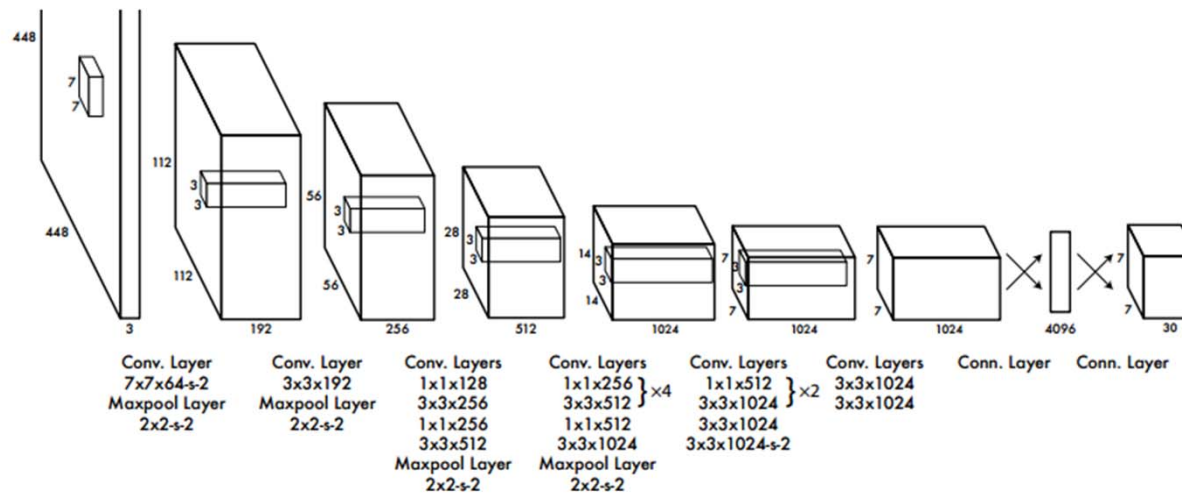


ED I2S : M

19 avril 2019 - Jean-Christophe Burie

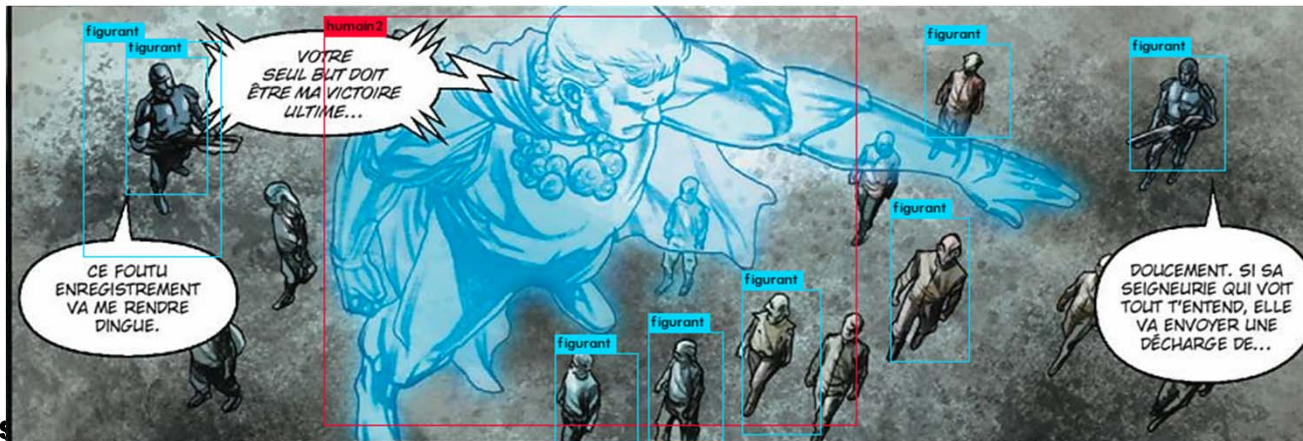
Etape 2 : Détection de personnages

- + Architecture du réseau
 - > Yolo v2



Etape 2 : Détection de personnages

+ Résultats

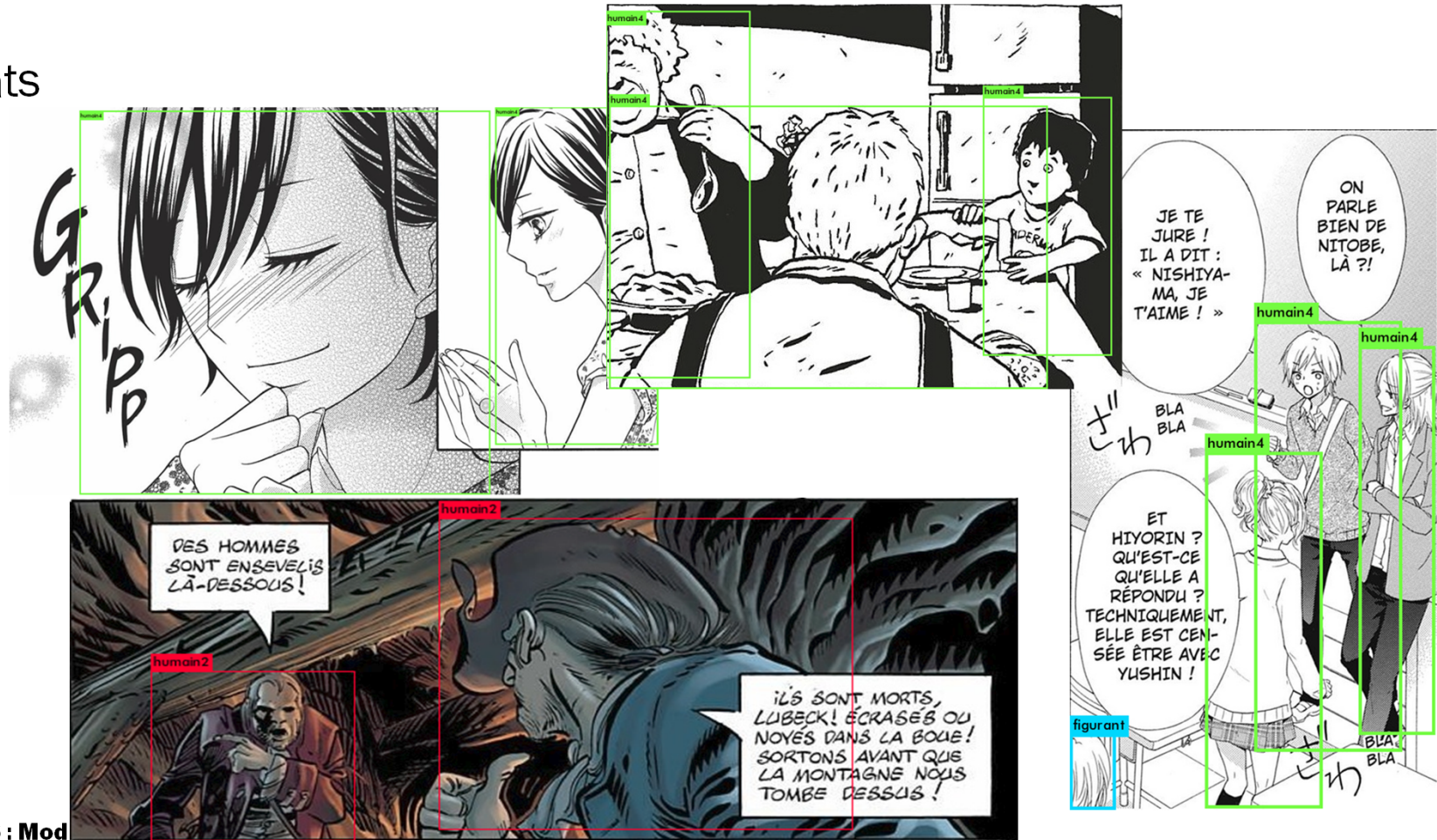


ED 129

19 avril 2019 - Jean-Christophe Burie

Etape 2 : Détection de personnages

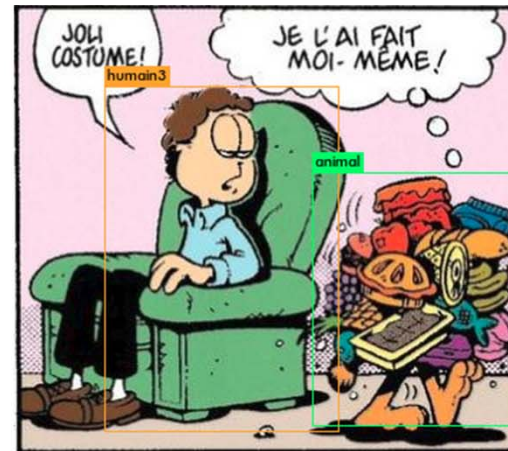
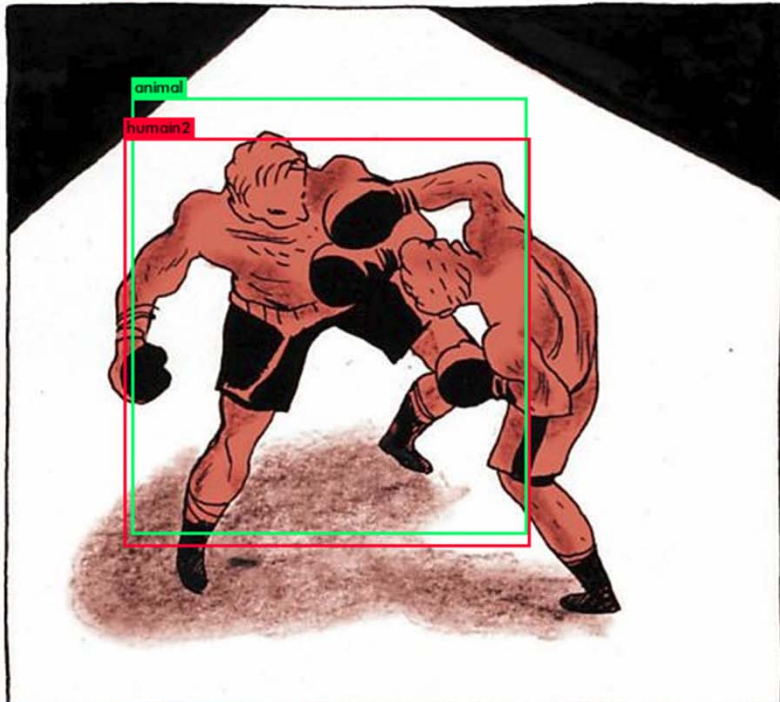
+ Résultats



ED I2S : Mod

19 avril 2019 - Jean-Christophe Burie

Etape 2 : Détection de personnages



19 avril 2019 - Jean-Christophe Burie

Etape 2 : Détection de personnages

+ Conclusions

- > Apprentissages profonds fonctionnent
- > Mais nécessite des données d'apprentissage

+ Travaux actuels

- > Apprentissage avec peu de données annotées

COMBINAISON D'APPROCHES

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

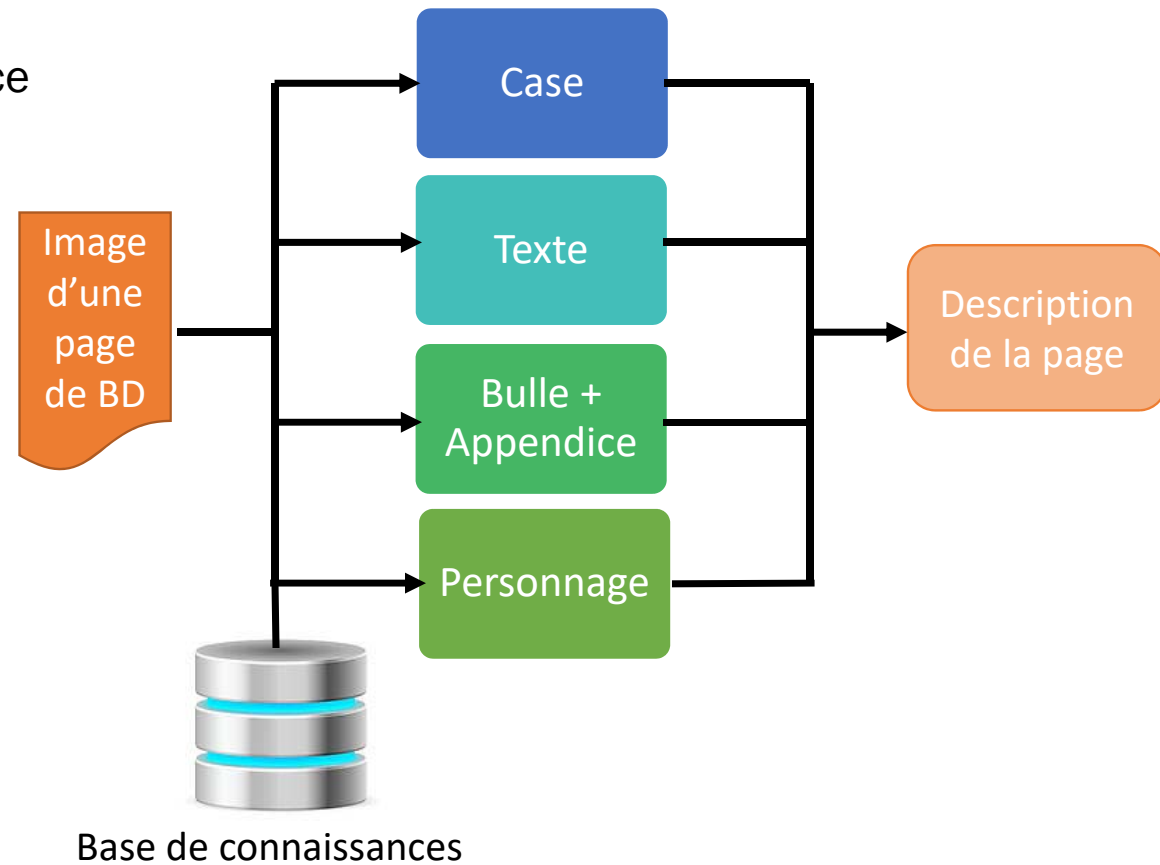
- + Méthodes ne sont pas parfaites
 - > Méthode classiques
 - > Méthode à base de réseaux de neurones

- + Combinaison d'approches
 - > Extraction conduite par la connaissance

 - > Comic MTL: modèle multi-tâches pour l'analyse des BD

+ Principe

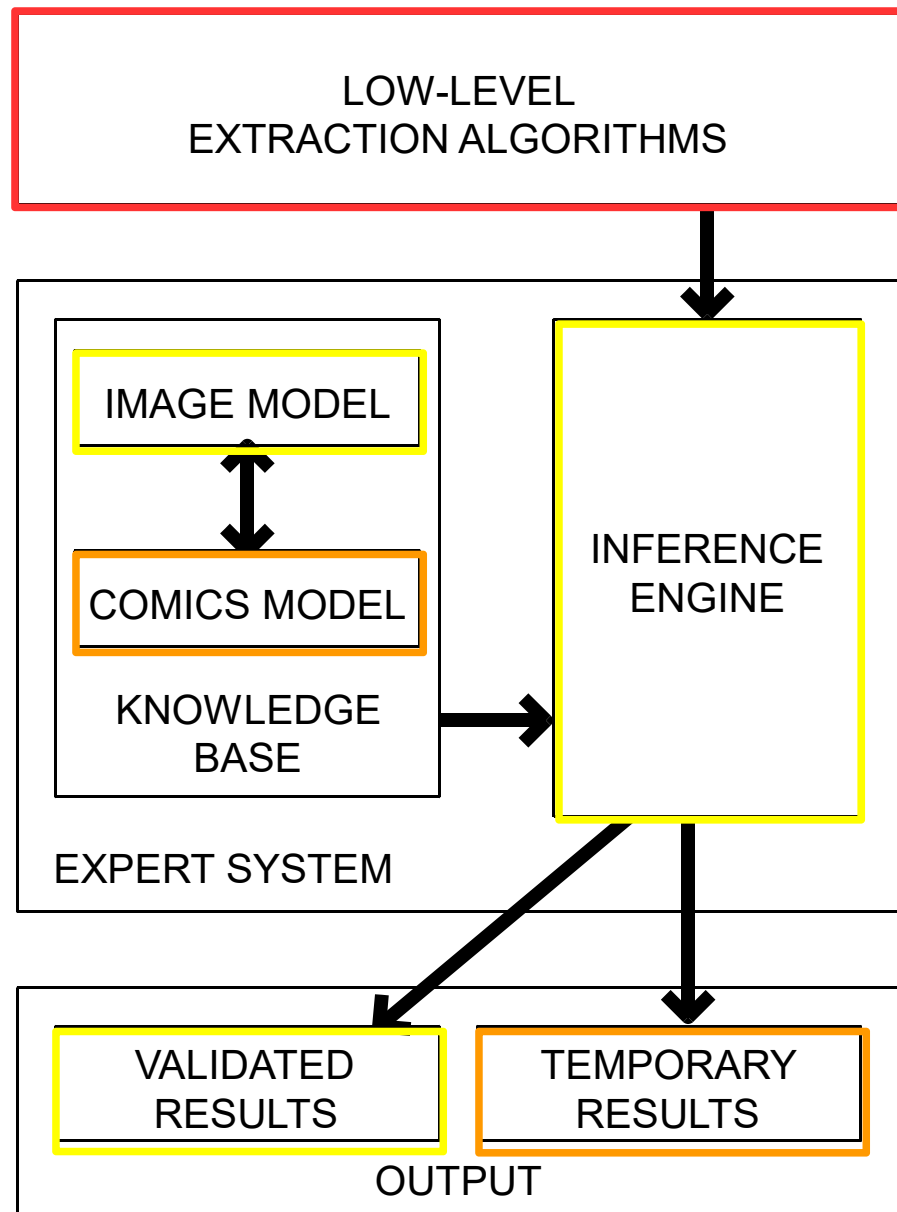
- > Modéliser la connaissance
- > Retrouver le contexte



Knowledge representation

Contributions

Knowledge-driven approach



- Rigaud's thesis
- Collaboration
- Guérin's thesis

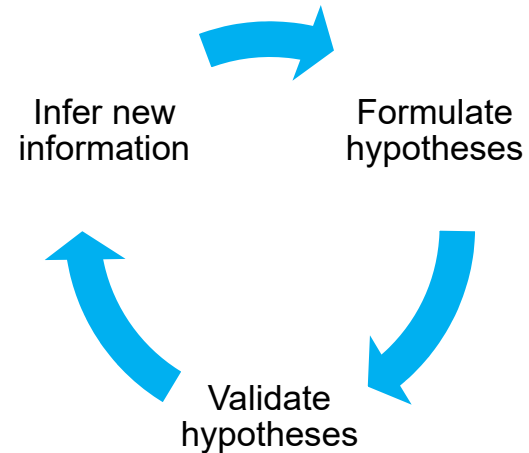
- Image model
 - Physical support
 - Regions of interest
- Comics model
 - Validations
 - A **panel P** is related to one page
 - A **balloon B** is related to one panel and may have a **tail Q**
 - A **character C** is related to one panel
 - A **text line T** is related to one balloon
 - Inferences
 - $B + Q + T \Rightarrow$ **speech balloon SB**
 - $SB + T \Rightarrow$ **speech text ST**
 - $SB + C \Rightarrow$ **speaking character SC**

Processing sequence



Contributions

Knowledge-driven approach

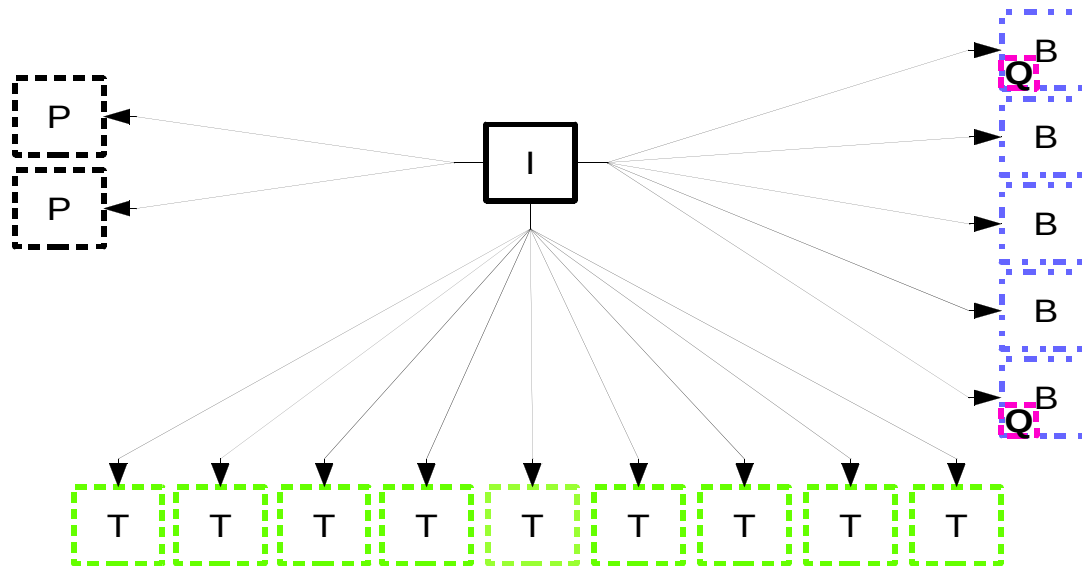
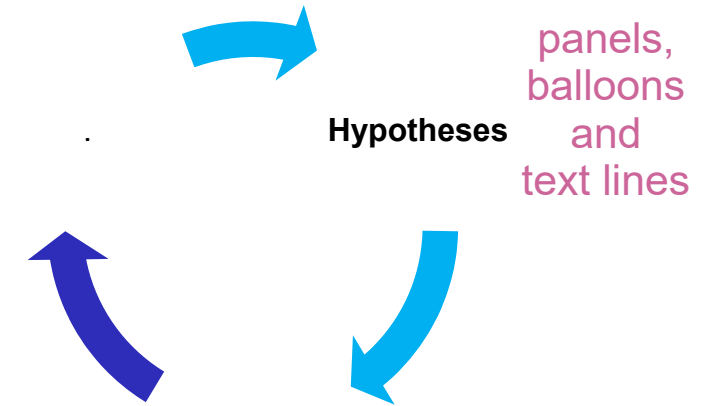


- Iteration 1
 - Step 1: hypotheses of **simple element** positions
 - Step 2: validation of the positions
 - Step 3: inference a new information
- Iteration 2
 - Step 1: hypotheses of more **complex elements**
 - Step 2: validation of the positions
 - Step 3: inference a new information

Processing sequence

Contributions

Knowledge-driven approach



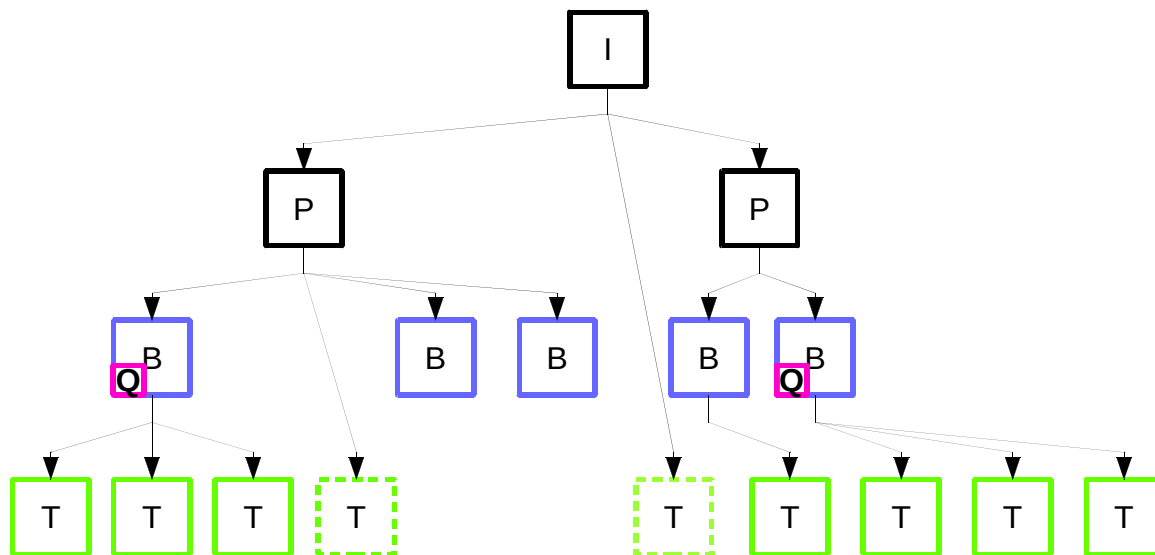
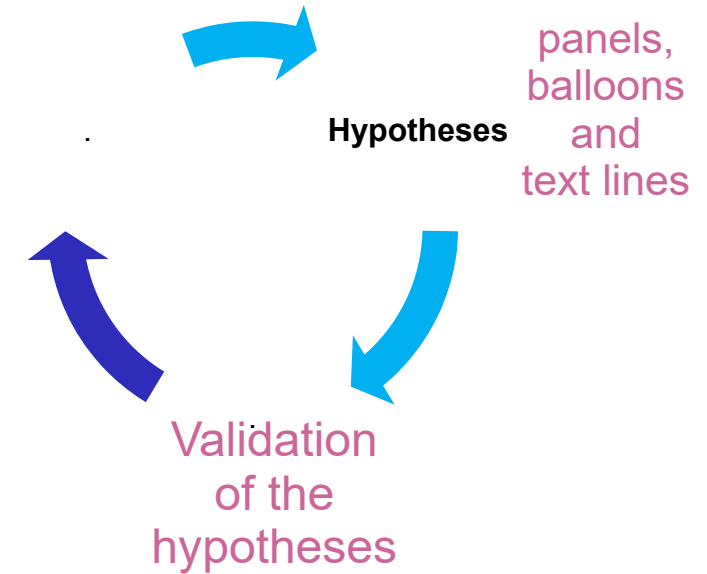
Non valid	Valid	
P	P	Panel
B	B	Balloon
Q	Q	Tail
T	T	Text

Processing sequence



Contributions

Knowledge-driven approach



Non valid	Valid	
P	P	Panel
B	B	Balloon
Q	Q	Tail
T	T	Text

Processing sequence



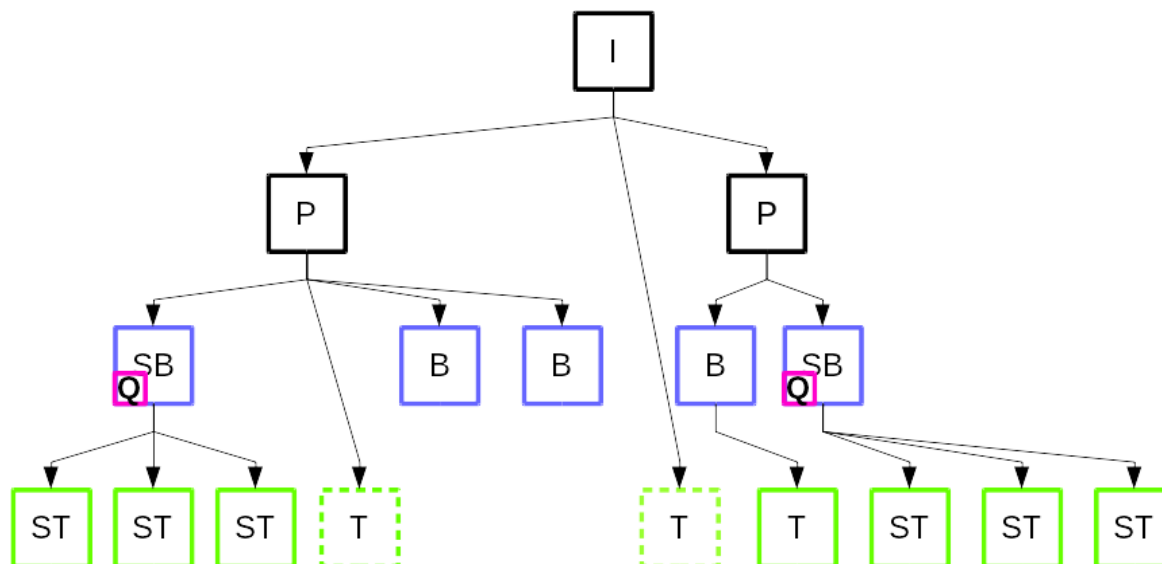
Contributions

Knowledge-driven approach

Inferences of specific types

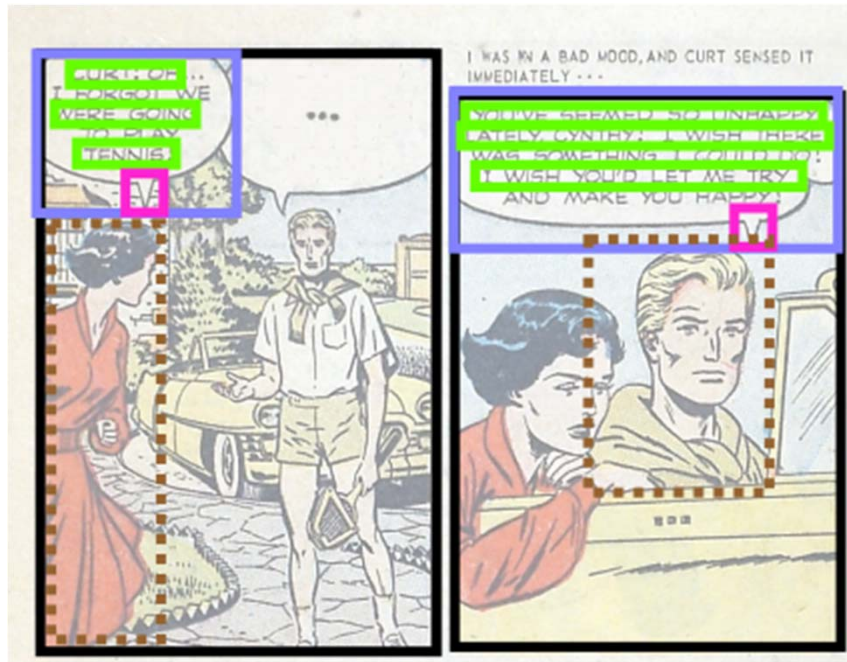
Hypotheses panels, balloons and text lines

Validation of the hypotheses



Non valid	Valid	
P	P	Panel
SB	SB	Speech balloon
Q	Q	Tail
ST	ST	Speech text

Processing sequence



Contributions

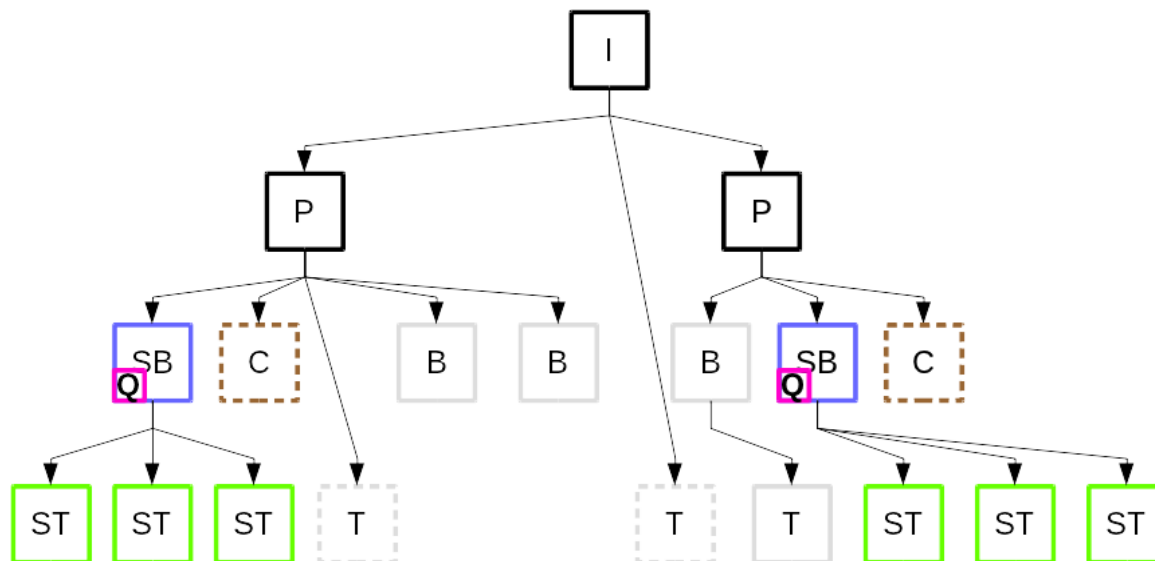
Knowledge-driven approach

Inferences of specific types

Hypotheses

Comic characters

Validation of the hypotheses



Non valid

Valid

C

C

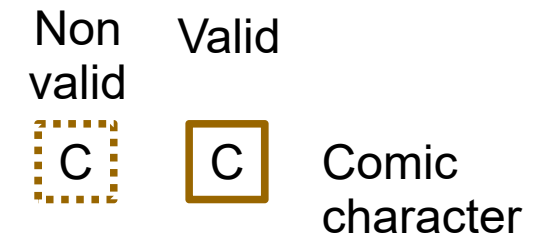
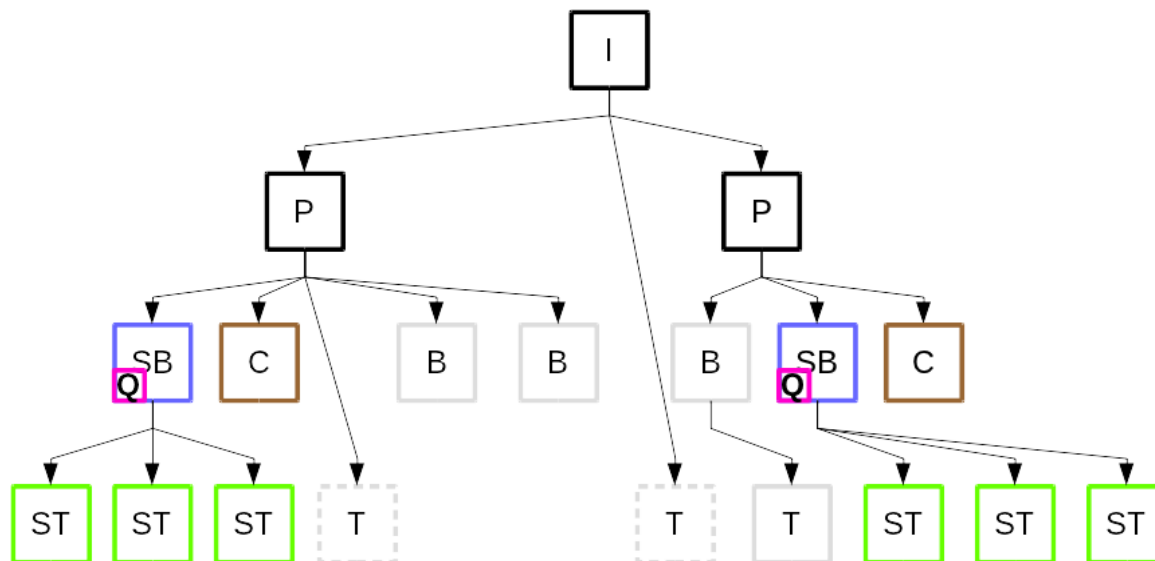
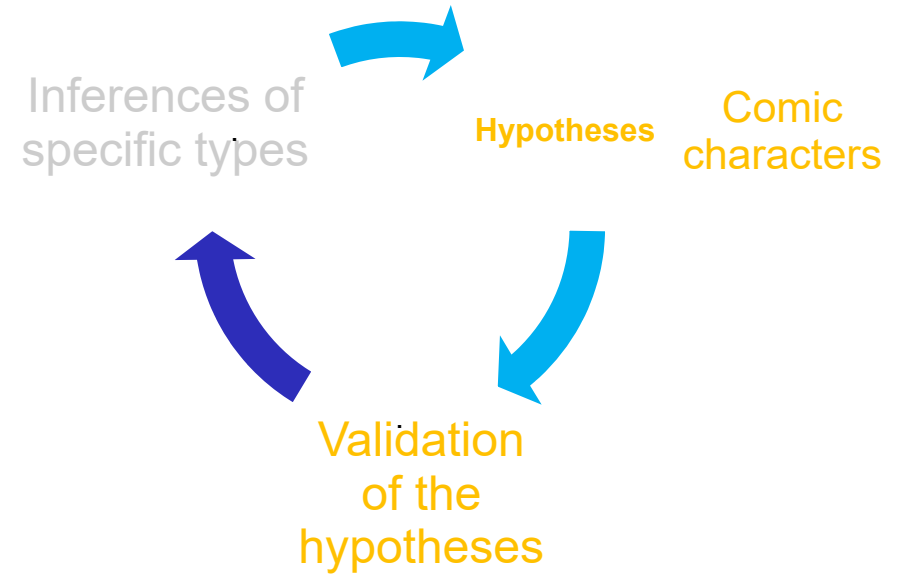
Comic character

Processing sequence



Contributions

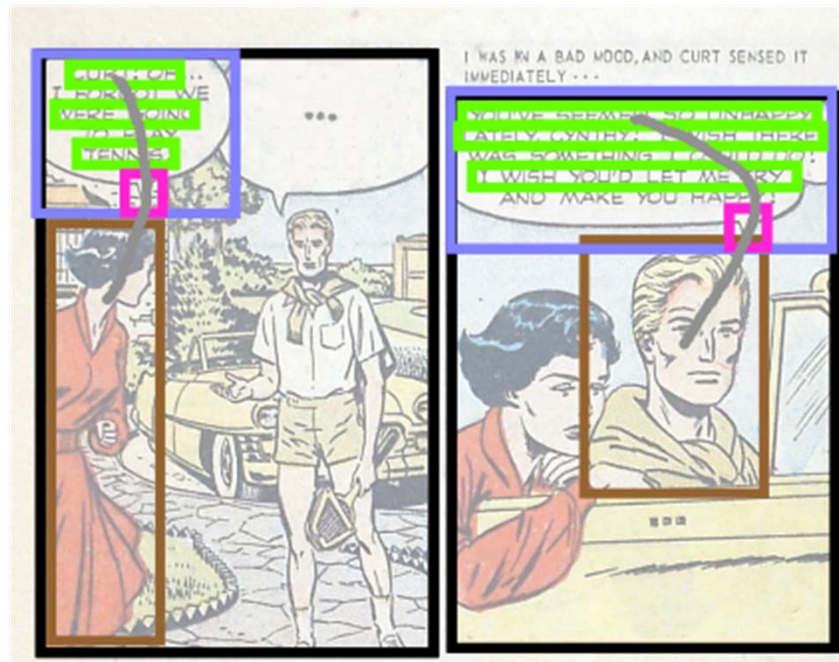
Knowledge-driven approach



Processing sequence

Contributions

Knowledge-driven approach

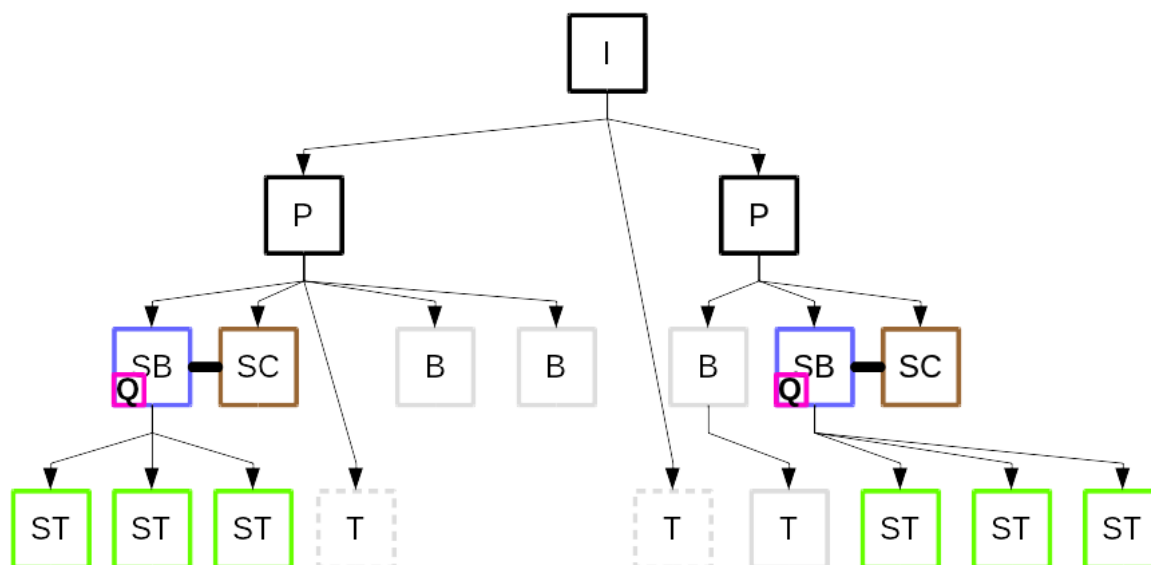


Inferences of specific types + semantic links

Hypotheses

Comic characters

Validation of the hypotheses



Non valid

Valid

SC

SC

Speaking comic character

Evaluations

Experiments

B_p = predicted region

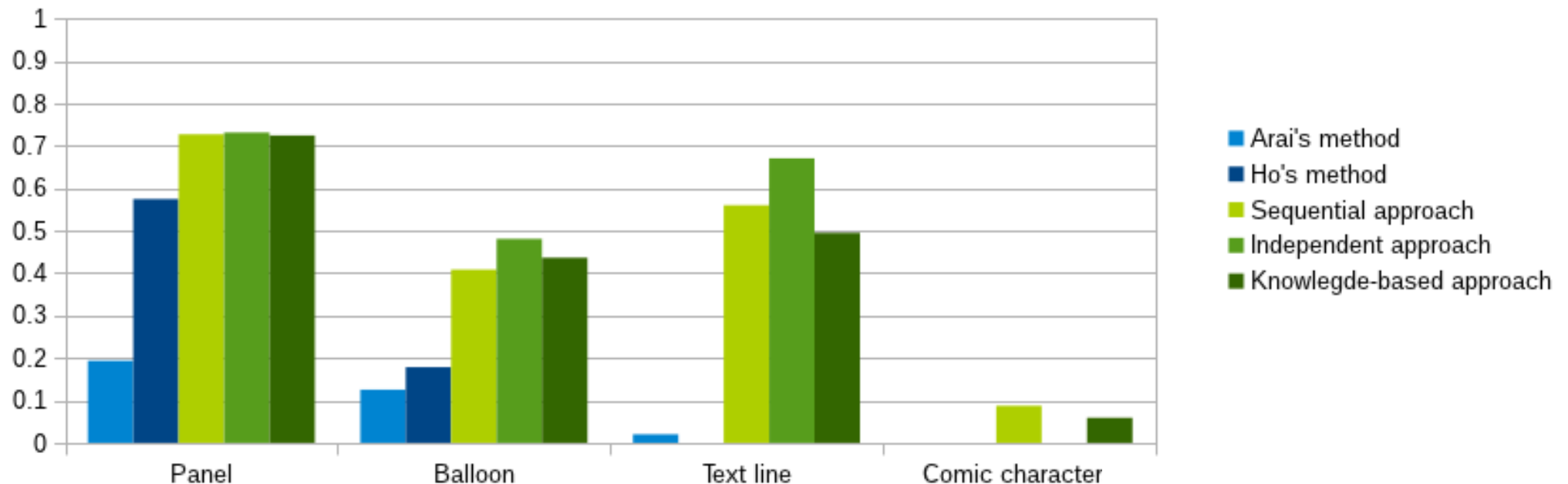
B_{gt} = ground truth region

B_p is valid if $a_0 > 0.5$

$$a_0 = \frac{\text{area}(B_p \cup B_{gt})}{\text{area}(B_p \cap B_{gt})}$$

Number of element in the dataset	
Panel	850
Balloon	1092
Text line	4691
Comic character	1550

Extraction results on the eBDtheque dataset (F-score)



Comic MTL: modèle multi-taches pour l'analyse des BD

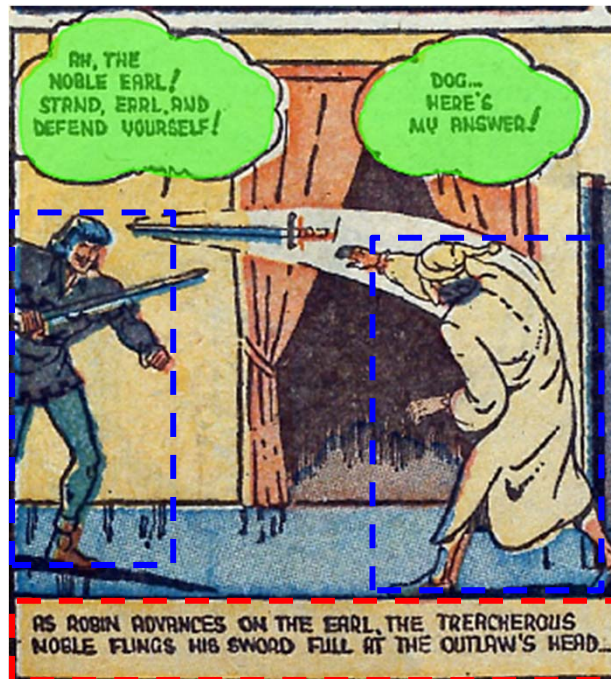


Comic MTL: multi-task model for comic book image analysis

Introduction

Comic book image analysis

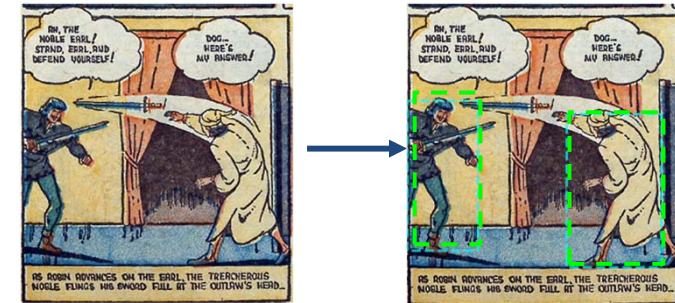
- Elements
 - > Balloon, text, panel, narrative box, character and other scene objects
- Tasks
 - > Elements related tasks (detection, segmentation, relationship analysis ...)
 - > Other tasks (style identification, story/scene understanding, colorization ...)



Introduction

Comic book image analysis

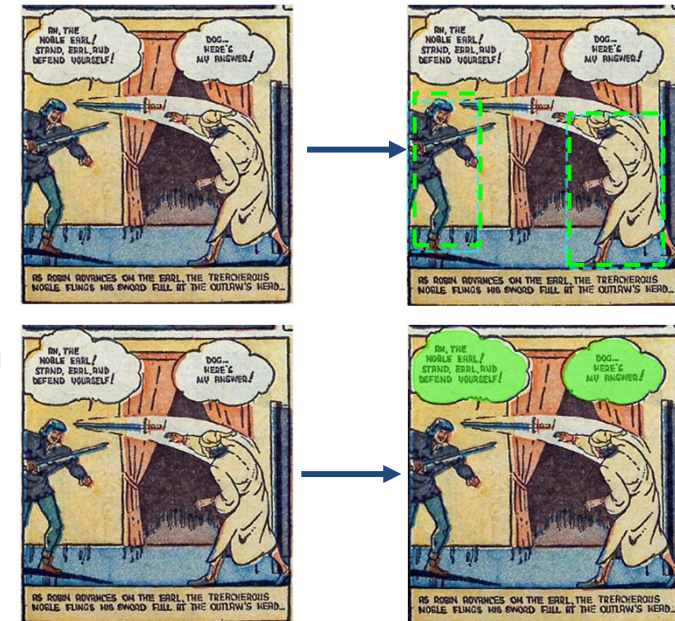
- **Detection task**
 - > Panels, characters, objects, narrative text boxes
 - > Multiple detection algorithms
 - > Or one deep learning detection model



Introduction

Comic book image analysis

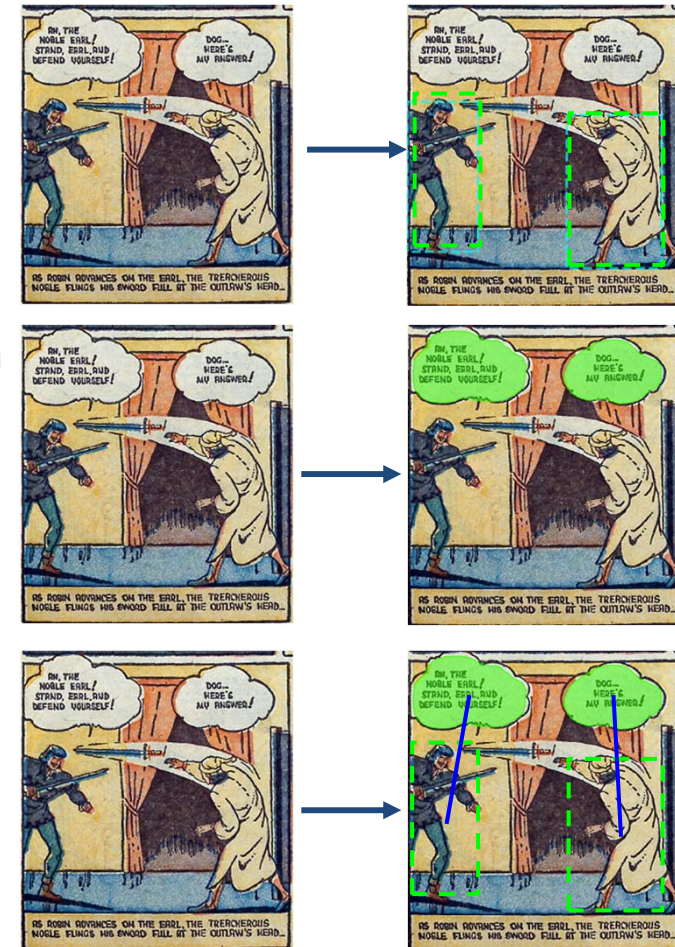
- Detection task
 - > Panels, characters, objects, narrative text boxes
 - > Multiple detection algorithms
 - > Or one deep learning detection model
- Segmentation task
 - > Balloons
 - > One segmentation algorithm or one segmentation deep learning model



Introduction

Comic book image analysis

- Detection task
 - > Panels, characters, objects, narrative text boxes
 - > Multiple detection algorithms
 - > Or one deep learning detection model
- Segmentation task
 - > Balloons
 - > One segmentation algorithm or one segmentation model
- Relationship analysis task
 - > Between detected elements
 - > E.g. : read before, said by, thought by, addressed to...
 - Balloon-character (balloon and its speakers)
 - > One or multiple algorithms/models

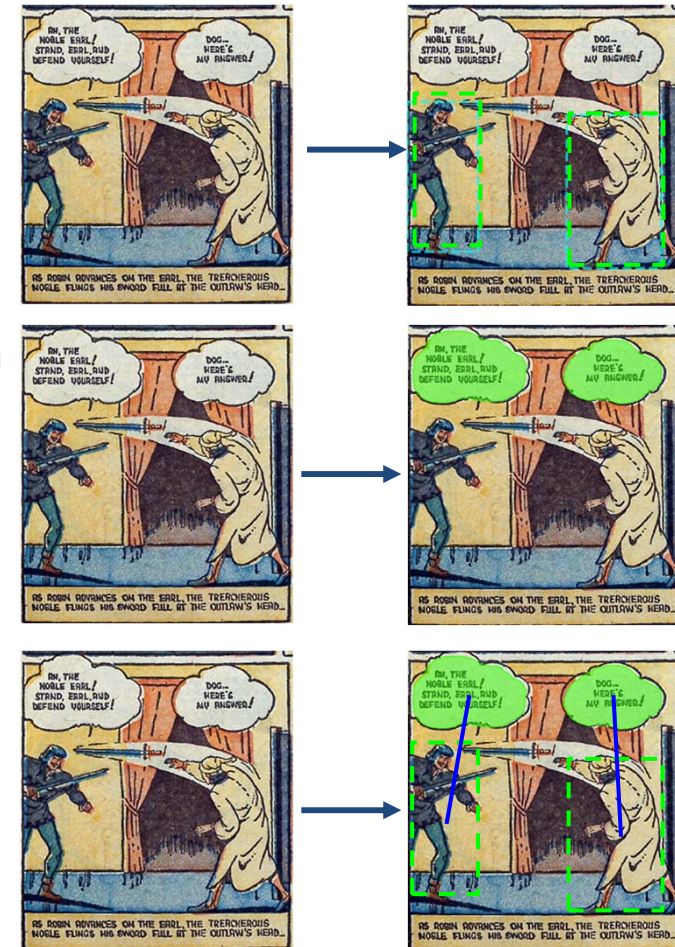


Introduction

Comic book image analysis

- Detection task
 - > Panels, characters, objects, narrative text boxes
 - > Multiple detection algorithms
 - > Or one deep learning detection model
- Segmentation task
 - > Balloons
 - > One segmentation algorithm or one segmentation model
- Relationship analysis task
 - > Between elements detections
 - > E.g. : read before, said by, thought by, addressed to...
 - Balloon-character (balloon and its speakers)
 - > One or multiple algorithms/models

=> *Multiple algorithms or models*

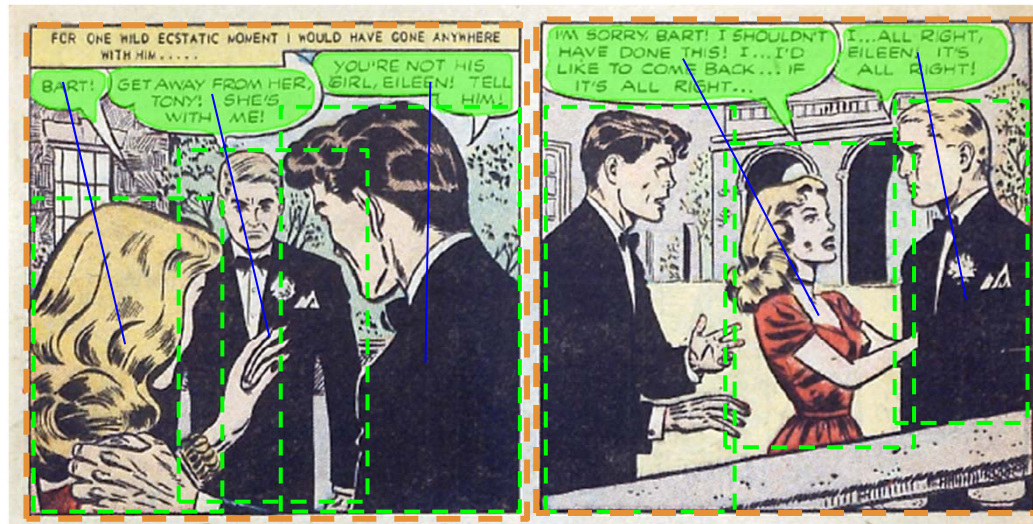


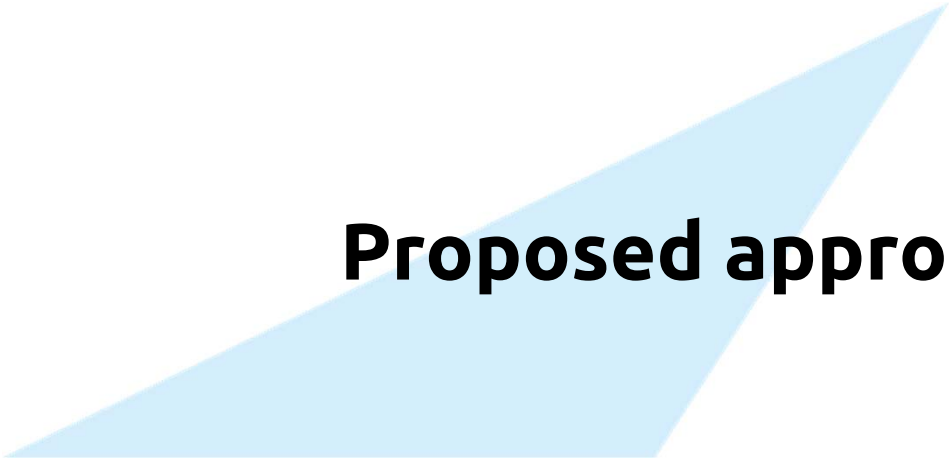
Introduction

Comic book image analysis

=> *Multiple algorithms or models*

- Comic MTL
 - > One CNN model for multiple tasks
 - > Detection, segmentation, relationship analysis
 - > Panels, characters, faces, narrative text boxes, balloons, associations balloon-character (or balloon-face)



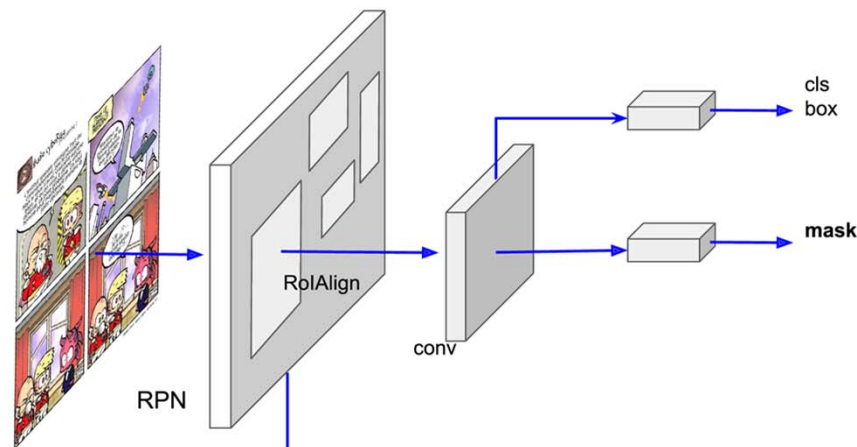


Proposed approach

Proposed approach

Comic MTL : Mask R-CNN

- Mask R-CNN model
 - > detection branch
 - > segmentation branch
 - > learn in parallel, but need segmentation masks for all classes in the ground-truth

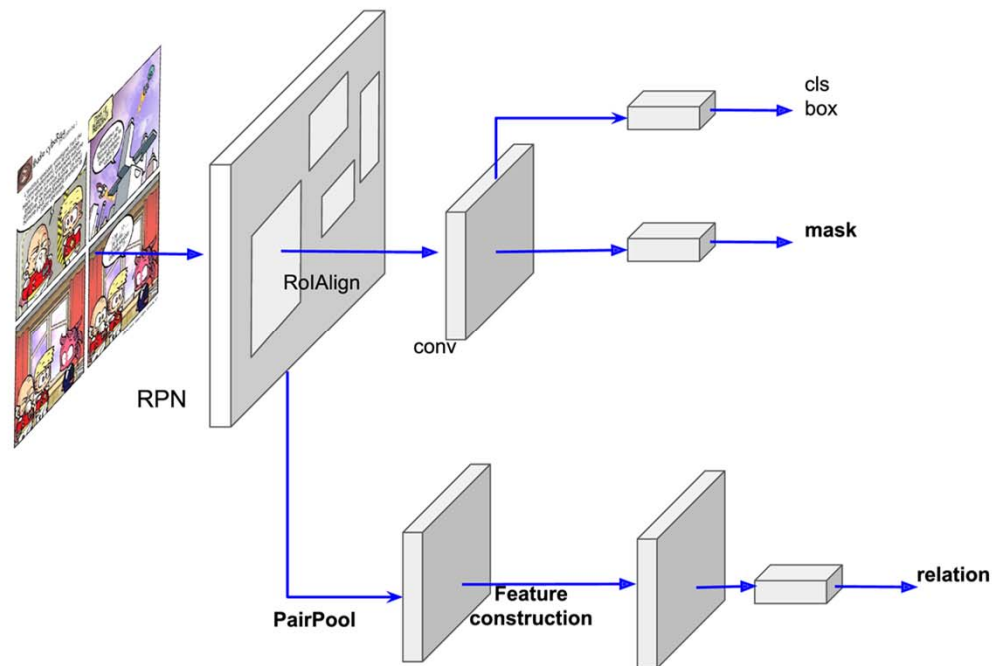


A set of candidate boxes
proposal

Proposed approach

Comic MTL Overview

- Mask R-CNN
- Simultaneously learning from bounding boxes and segmentation masks
- Relationship analysis for balloons and characters



Proposed approach

Comic MTL : detection and segmentation

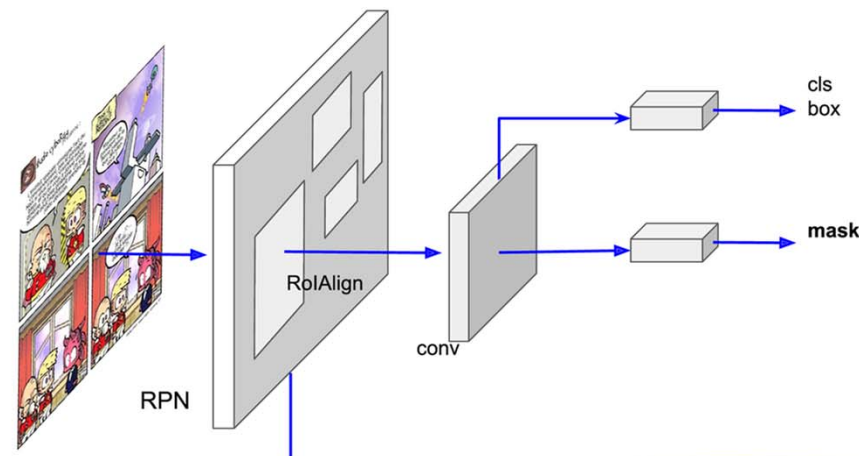
- Mask R-CNN

$$L = L_{cls} + L_{box} + L_{mask}$$

$$L_{mask} = -\frac{1}{m^2} \sum_{i,j} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^k)]$$

- Comic MTL: simultaneously learning from bounding boxes and segmentation masks

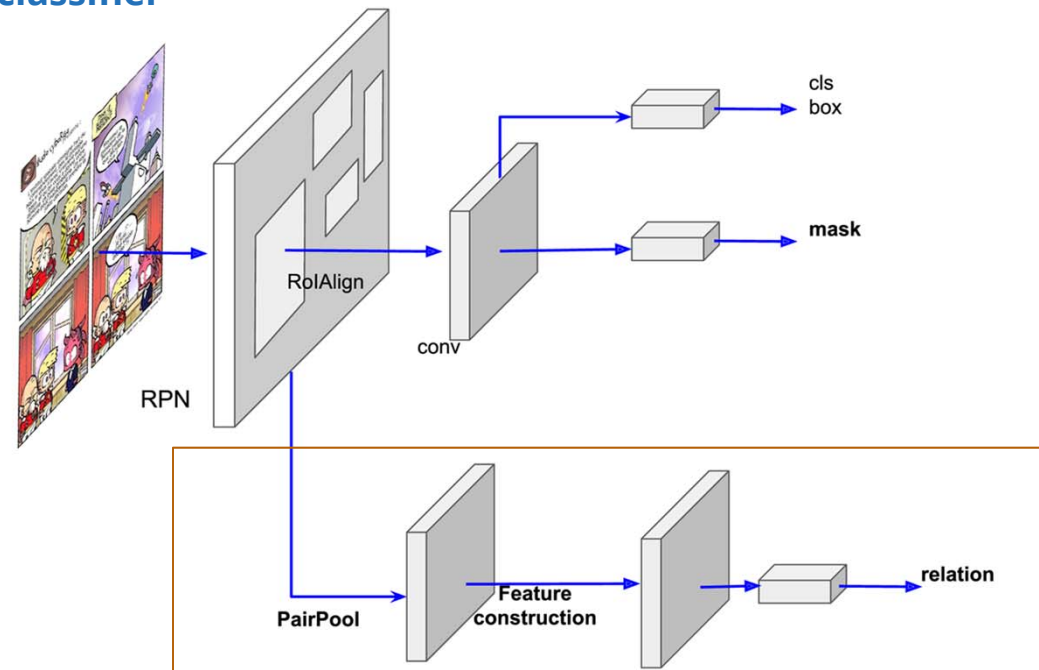
$$L_{mask} = -\frac{I^k}{m^2} \sum_{i,j} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^k)]$$



Proposed approach

Comic MTL : relation branch

- Mask R-CNN
- Simultaneously learning from bounding boxes and segmentation masks
- **Relationship analysis for balloons and characters**
 - > Check if a pair balloon-character “has a link”
 - The character is the speaker of the balloon
 - > Additional relation branch : **a classifier**

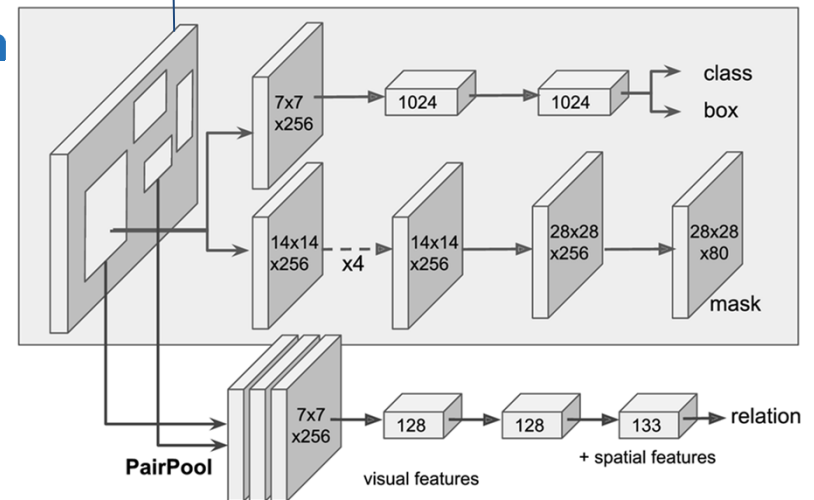


Proposed approach

Comic MTL : head architecture

- Relationship analysis for balloons and characters
 - > Additional relation branch
- Balloon-character pairs combination**
 - **PairPool**
 - > Best candidates boxes
 - > Positive pairs vs Negative pairs = 1:1
 - > Pad 0 at the end to have N pairs (N is fixed)

A set of candidate boxes proposal



Proposed approach

Comic MTL : head architecture

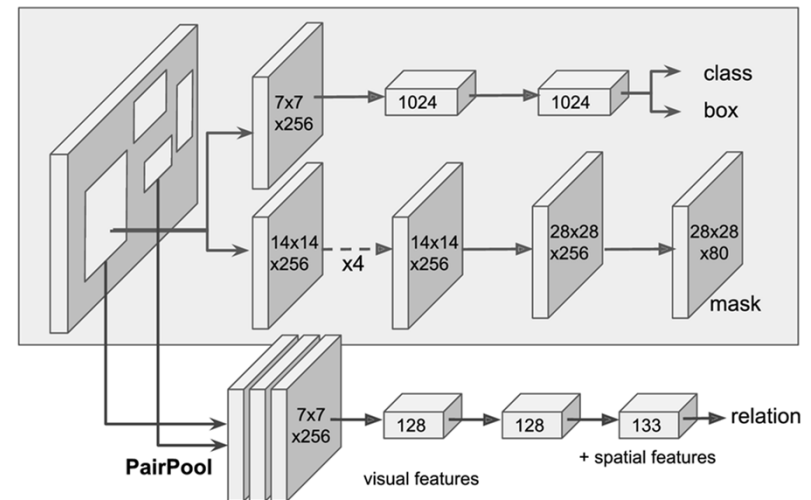
- Relationship analysis for balloons and characters
 - > Additional relation branch
- Balloon-character pairs combination - PairPool
- Features construction**
 - > Visual features : union box
 - > Spatial features :

$$\left[\frac{x_b - x_c}{w_b}, \frac{y_b - y_c}{y_b}, \frac{x_b - x_c}{w_c}, \frac{y_b - y_c}{y_c}, \frac{b \cap c}{b \cup c} \right]$$

where 2 boxes (balloon and character):

$$b = [x_b, y_b, w_b, h_b]$$

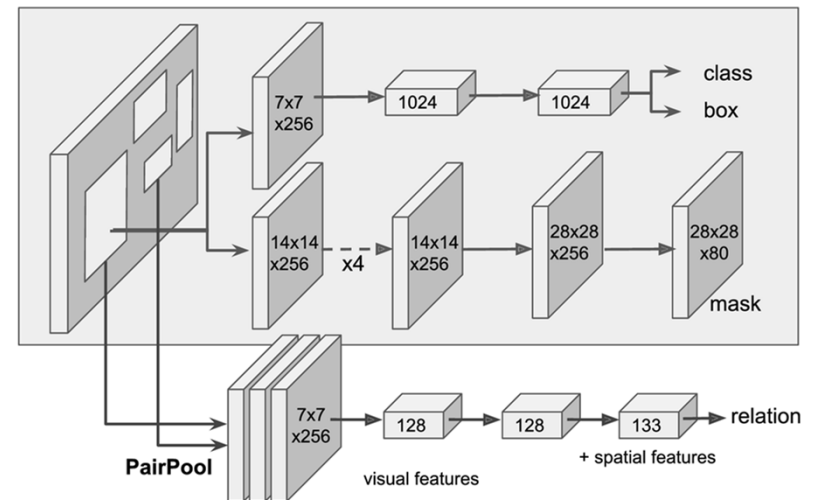
$$c = [x_c, y_c, w_c, h_c]$$



Proposed approach

Comic MTL : head architecture

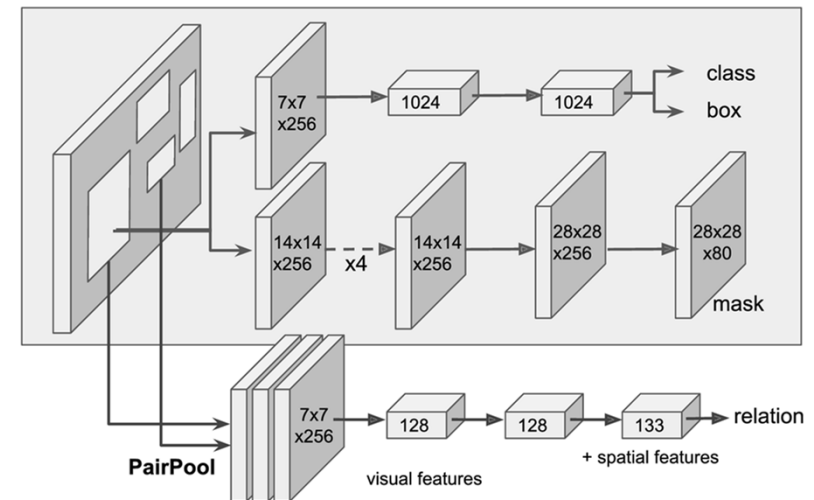
- Relationship analysis for balloons and characters
 - > Additional relation branch
- Balloon-character pairs combination - PairPool
- Features construction
 - > Visual features : union box
 - > Spatial features :
- **Training vs Inference**
 - > Parallel training for 3 branches
 - > Inference: take the output of detection branch to feed into relation branch



Proposed approach

Comic MTL : head architecture

- Relationship analysis for balloons and characters
 - > Additional relation branch
- Balloon-character pairs combination - PairPool
- Features construction
 - > Visual features : union box
 - > Spatial features :
- Training vs Inference
 - > Parallel training for 3 branches
 - > Inference: take the output of detection branch to feed into relation branch
- Backbone**
 - > ResNet 50 + FPN



Proposed approach

Comic MTL : summary

- **Comic MTL**
 - > Based on **Mask R-CNN**
 - > Simultaneously learning from **bounding boxes** and **segmentation masks**
 - > **Additional branch** for relationship analysis
- **Multi-tasks:**
 - Detection task
 - > Panels, comic characters, face, narrative boxes
 - Segmentation task
 - > Balloons
 - Relationship analysis
 - > Balloons - characters or balloons - faces

Evaluation

Evaluation

Experimental protocol

- Datasets
 - > eBDtheque
 - the only dataset has balloon segmentation masks and associations between balloons and characters
 - 100 comic book images
 - 850 panels
 - 1550 comics characters
 - 1092 balloons (including narrative text boxes)
 - 814 links balloon-character
- Evaluation
 - > 10-fold cross-validation on eBDtheque
 - > (90 train images and 10 test images)
 - > Transfer learning : using pre-trained model from ImageNet

Evaluation

Results : Segmentation

- Comic MTL: Balloon segmentation
 - > In comparison with traditional algorithms
 - > In comparison with Mask R-CNN
 - > The same setting for Comic MTL and Mask R-CNN

Method	Recall	Precision	F-Measure
Arai [3]	18.70	23.14	20.69
Ho [11]	14.78	32.37	20.30
Rigaud [27]	69.81	32.83	44.66
Rigaud [25]	62.92	62.27	63.59
Mask R-CNN [10]	75.31	92.42	82.99
Comic MTL	74.94	92.77	82.91

Evaluation

Results : Detection

- Comic MTL: Detection task
 - > In comparison with existing works

Characters detection

Method	Characters recall	Characters precision	F1
Arai [3]	-	-	-
Rigaud [29]	-	-	-
Rigaud [26]	21.6	40.5	28.2
Ogawa [19]	42.2	58.0	48.8
Comic MTL	62.17	71.79	66.63

(*) The model in Ogawa [19] is trained on another dataset (Manga109)

Evaluation

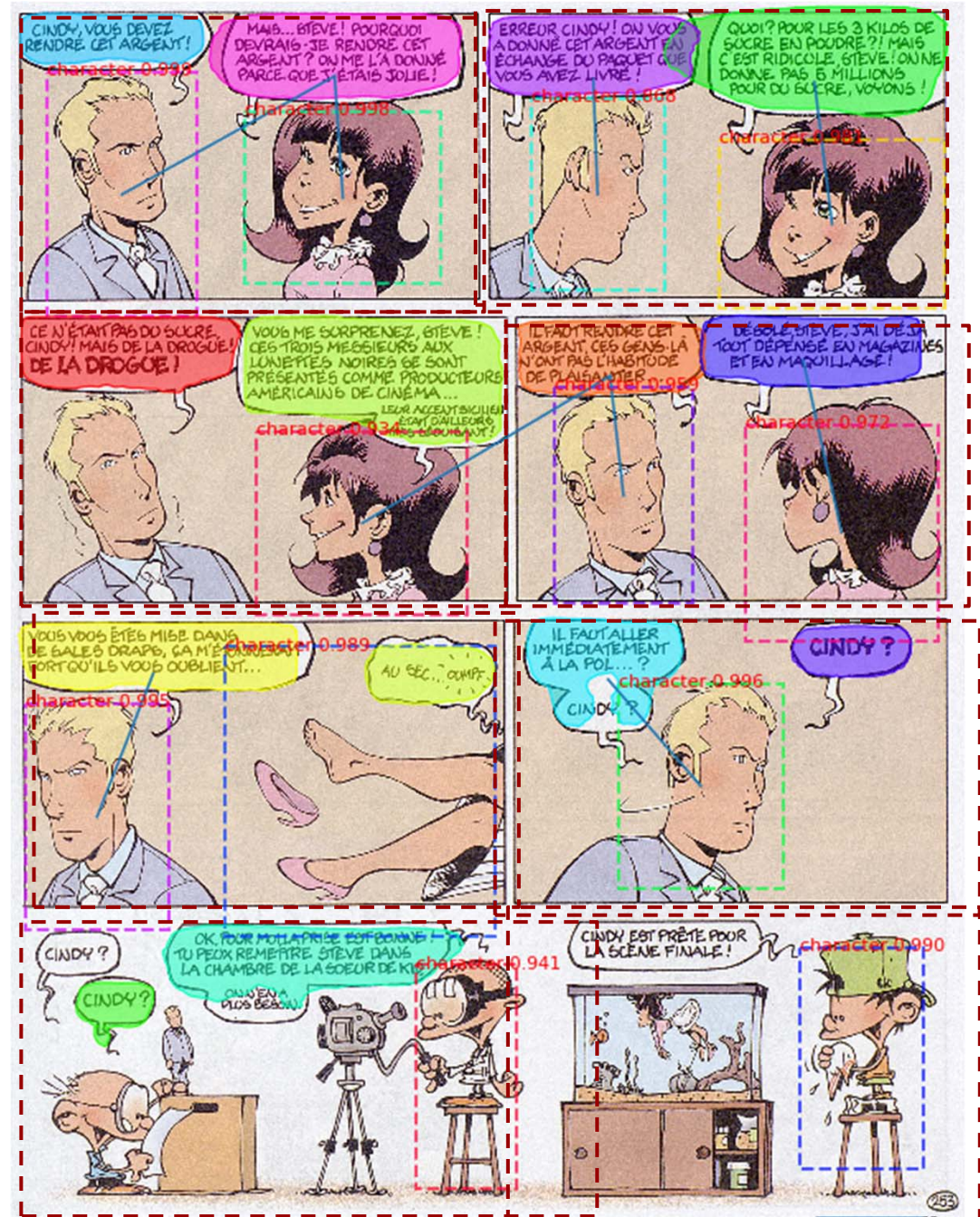
Results : Relationship analysis

- Relationship analysis
 - > In comparison with the work of Rigaud (2015)
 - > Setting 1: no annotations for characters & balloons
 - > Setting 2: A priori bounding boxes for character, masks for balloon and balloon tails

Method	Recall	Precision	F1
Setting 1 [23]	-	18.01	-
Setting 2 [23]	-	93.32	-
Comic MTL	28.32	52.88	36.89

Evaluation Result examples

- balloon masks
- bounding boxes of characters
- the lines that connect a balloon and a character.
- 2 balloons not detected (at the bottom).
- 2 characters not detected (second row and last row).
- 7 good links detected, and two bad links detected.



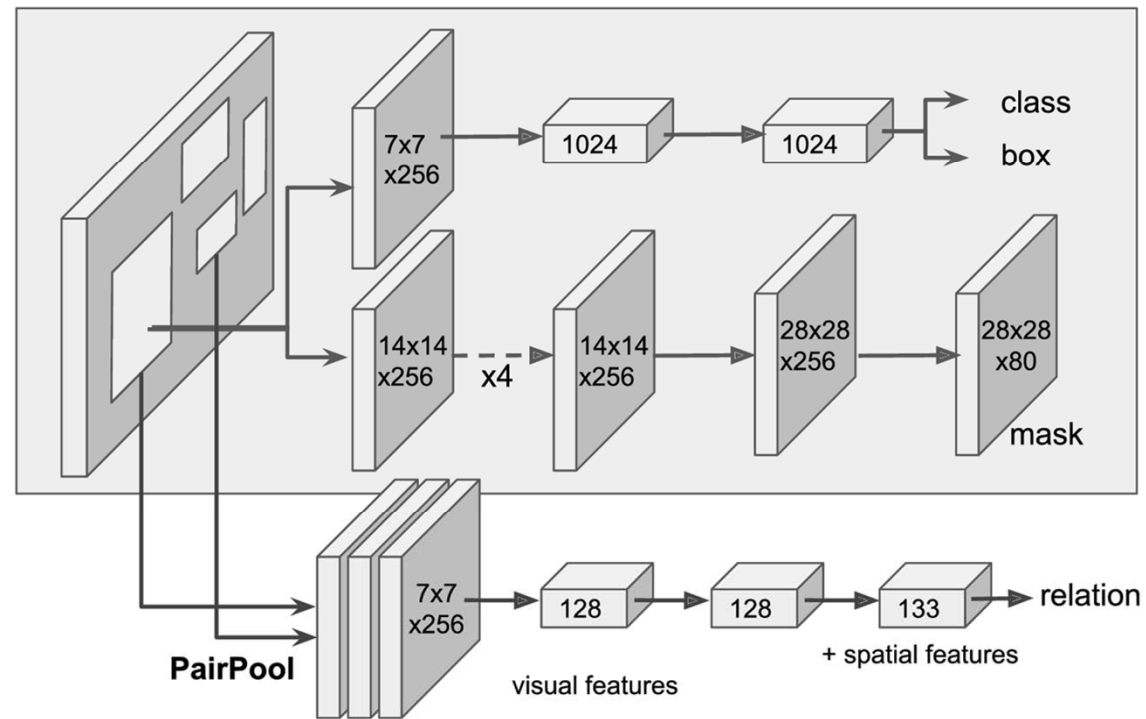
Conclusions

Conclusion

- Multiple tasks on one CNN model
 - > characters detection, panels and balloons segmentation, and balloon-character association analysis for comic book images.
- State-of-the-art performance on detection and segmentation tasks
- Promising results on relationship analysis
 - > Perspective:
 - balloon and its speaker (character) should be in the same panel
 - learning the direction of the balloon tail

Proposed approach

Head architecture



Evaluation

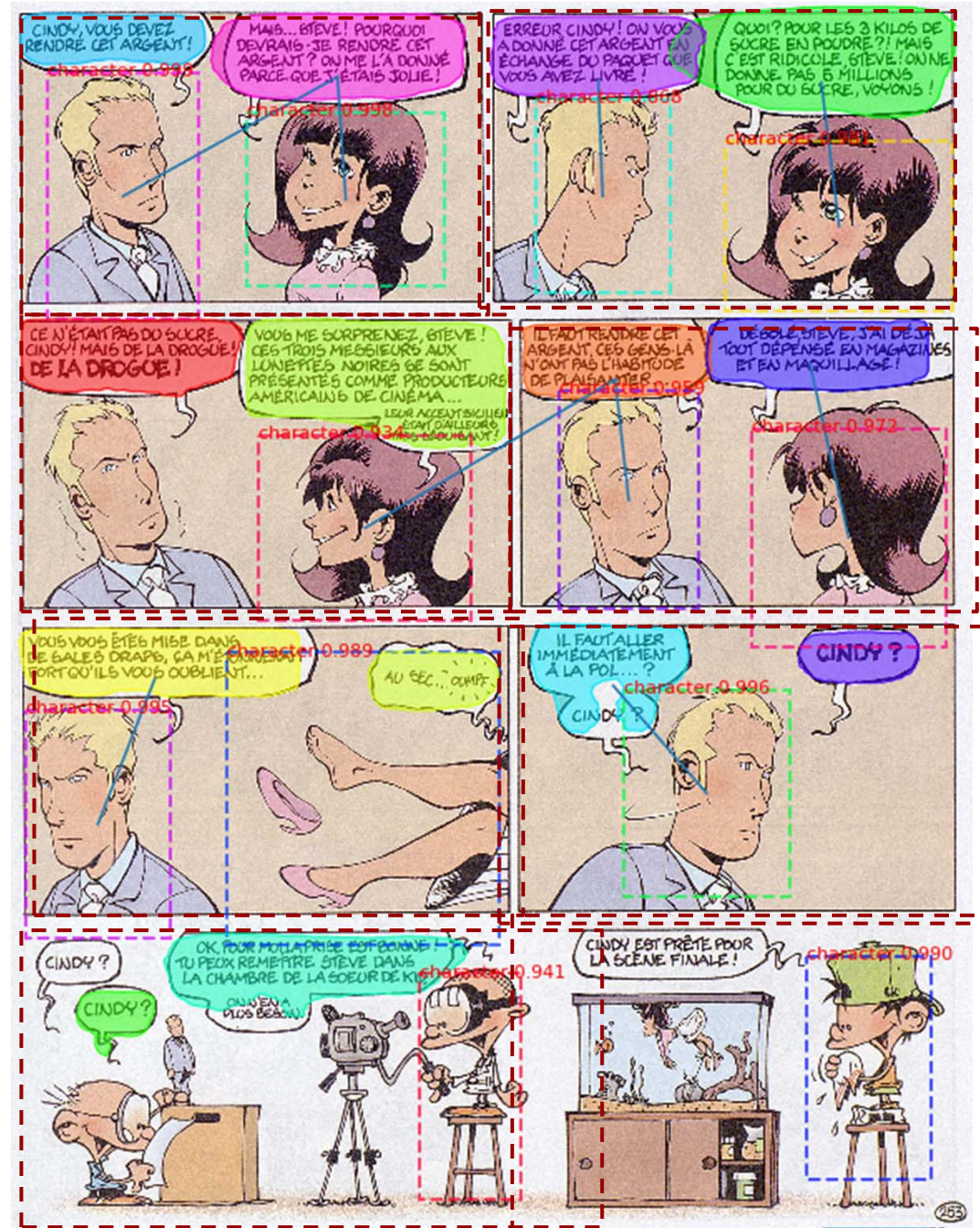
Results : Process time evaluation

- The Comic MTL model takes 30\% less time as compared with the total time of Mask R-CNN and Faster R-CNN models

	Faster R-CNN	Mask R-CNN	Comic MTL	Mask R-CNN + Faster R-CNN
Training	80 mins	95 mins	120 mins	175 mins
Inference	0.177s	0.210s	0.273s	0.387s

Evaluation Result examples

- balloon masks
- bounding boxes of characters
- the lines that connect a balloon and a character.
- 2 balloons not detected (at the bottom).
- 2 characters not detected (second row and last row).
- 7 good links detected, and two bad links detected.



Quelques résultats récents

- + Exemples non diffusable pour des raisons de propriétés intellectuelles

CONCLUSIONS & PERSPECTIVES

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

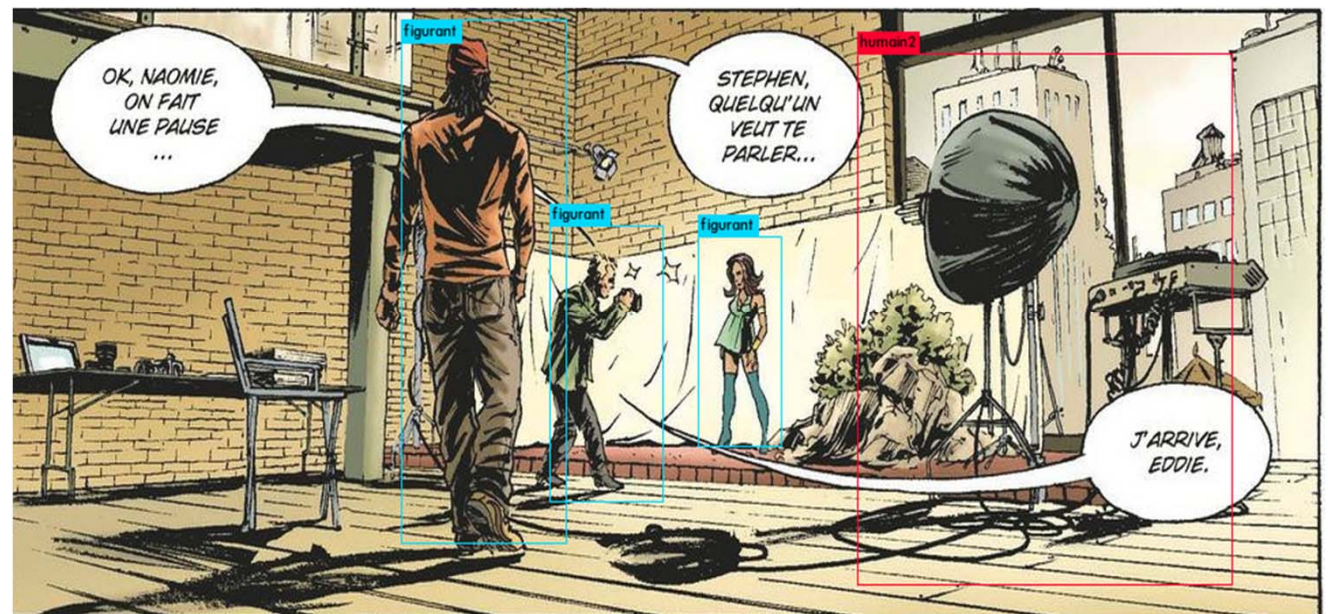
- + Les pages de bandes dessinées sont des « objets » complexes
- + Difficultés d'avoir une méthode générique
 - > Combinaisons d'approches
- + Travail d'équipe :
 - > 4 permanents > 1 IGR
 - > 3 doctorants > 4 post-docs
- + Création d'un laboratoire commun



Sequential [Art Image](#) Laboratory

+ Aller plus loin dans l'analyse des cases

- > Contenu : objets, décors, actions, ...
- > Analyse du texte..
- > Relations sémantiques
- > ...



+ Nouvelles formes de BD

+ Exemples:

> <http://toy-planes.com/index.html>

> <https://www.webtoons.com/en/>

BIBLIOGRAPHIE

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Revues internationales

- > Van Nhu Nguyen, Christophe Rigaud, Jean-Christophe Burie, Digital comics image indexing based on deep learning, JI : Journal of Imaging, Special Issue Image Based Information Retrieval from the Web, volume 4, issue n° 7, article 89, 2018 <https://doi.org/10.3390/jimaging4070089>
- > Thanh Nam LE, Muhammad Muzzamil Luqman, Anjan DUTTA, Pierre HEROUX, Christophe RIGAUD, Clément GUERIN, Pasquale FOGGIA, Jean-Christophe BURIE, Jean-Marc OGIER, Josep LLADOS, Sebastien ADAM, SSGCI: Subgraph Spotting in Graph Representations of Comic book Images, PRL : Pattern Recognition Letters, Volume 112, September 2018, Pages 118-124. <https://doi.org/10.1016/j.patrec.2018.06.017>
- > Christophe Rigaud, Clément Guérin, Dimosthenis Karatzas, Jean-Christophe Burie, Jean-Marc Ogier, Knowledge-driven understanding of images in comic books, IJDAR : International Journal on Document Analysis and Recognition, Volume 18, Issue 3, September 2015, Pages 199 – 221 - DOI : 10.1007/s10032-015-0243-1

+ Chapitre de livres

- > Christophe Rigaud, Jean-Christophe Burie, Computer Vision Applied to Comic Book Images, dans Empirical Comics Research : Digital, Multimodal, and Cognitive Methods, Edited By Alexander Dunst, Jochen Laubrock, Janina Wildfeuer, New York: Routledge, 2018. ISBN : 9781351733885 <https://doi.org/10.4324/97813515185354>
- > Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Text-independent speech balloon segmentation for comics and manga, dans LNCS : Lecture Notes in Computer Science : Graphics Recognition: Current Trends and Challenges, Springer Heidelberg, Volume 9657, Pages 133- 147, 2017.
- > Hoang Nam Ho, Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Detecting recurring deformable objects: an approximate graph matching method for detecting characters in comics books, dans LNCS: Lecture Notes in Computer Science : Graphics Recognition: Current Trends and Challenges, Springer Heidelberg, Volume 8746, Pages 122- 134, 2014.
- > Christophe Rigaud, Dimosthenis Karatzas, Jean-Christophe Burie, Jean-Marc Ogier, Adaptive contour classification of comics speech balloons, dans LNCS : Lecture Notes in Computer Science : Graphics Recognition: Current Trends and Challenges, Springer Heidelberg, Volume 8746, Pages 53- 62, 2014.
- > Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie, Jean-Marc Ogier, Robust frame and text extraction from comic books, dans LNCS : Lecture Notes in Computer Science : Graphics Recognition, Springer Heidelberg, Volume 7423, Pages 129-138, 2013.

+ Conférences internationales

- > Srikanta Pal, Jean-Christophe Burie, Umapada Pal, Jean-Marc Ogier, Line-wise Text Identification in comic books: A Support Vector Machine-based Approach, IJCNN 2016, International Joint Conference on Neural Networks, Vancouver, Canada, 24-26 Juillet 2016.
- > Christophe Rigaud, Nam Le Thanh, Jean-Christophe Burie, Jean-Marc Ogier, Motoi Iwata, Ekki Imazu, Koichi Kise, Speech balloon and speaker association for comics and manga understanding, ICDAR 2015, 13th Int. Conference on Document Analysis and Recognition, Nancy, France, 24-26 Août 2015.
- > Thanh Nam Le, Muhammad Muzzamil Luqman, Jean-Christophe Burie, Jean-Marc Ogier, Content-based Comic Retrieval Using Multilayer Graph Representation and Frequent Graph Mining, ICDAR 2015, 13th International Conference on Document Analysis and Recognition, Nancy, France, 24-26 Août 2015
- > Clément Guérin, Christophe Rigaud, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Antoine Mercier, Jean-Marc Ogier, Arnaud Revel, **eBDtheque: a representative database of comics**, ICDAR 2013, 12th International Conference on Document Analysis and Recognition, Washington, Etats-Unis, 25-28 Août 2013.
- > Weihang Sun, Jean Christophe Burie, Jean-Marc Ogier, Koichi Kise, Specific Comic Character Detection Using Local Feature Matching, ICDAR 2013, 12th International Conference on Document Analysis and Recognition, Washington, Etats-Unis, 25-28 Août 2013.
- > Christophe Rigaud, Jean Christophe Burie, Jean-Marc Ogier, Dimosthenis Karatzas, Joost van de Weijer, An active contour model for speech balloon detection in comics, ICDAR 2013, 12th International Conference on Document Analysis and Recognition, Washington, Etats-Unis, 25-28 Août 2013.

+ Workshop internationaux

- > Nhu Van Nguyen, Christophe Rigaud, Jean-Christophe Burie, Multi-task Model for Comic Book Image Analysis, MANPU 2019, 3rd ,International Workshop on coMics ANalysis, Processing and Understanding in conjunction with MMM 2019, 25th International Conference on Multimedia Modeling, p. 637-649, Thessaloniki, Greece, January 8-11, 2019,
- > Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Segmentation-free speech text recognition for comic books, MANPU 2017, 2nd International Workshop on coMics ANalysis, Processing and Understanding, article n° 1, p 1:1-1:6, Kyoto, Japan, 10 Novembre 2017
- > Van Nguyen, Christophe Rigaud, Jean-Christophe Burie, Comic characters detection using deep learning, MANPU 2017, 2nd International Workshop on coMics ANalysis, Processing and Understanding, article n° 1, p 1:1-1:6, Kyoto, Japan, 10 Novembre 2017

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

+ Workshop internationaux (suite)

- > Clément Guérin, Jean-Christophe Burie, Jean-Marc Ogier, Detection of comic books twin pages with a non-overlapping stitching method, MANPU 2016, 1st Int. Workshop on coMics ANalysis, Processing and Understanding, Cancun, Mexique, 4 Décembre 2016, doi : 10.1145/3011549.3011550
- > Thanh Nam Le, M. Muzzamil Luqman, Jean-Christophe Burie, J.-M. Ogier, Retrieval of comic book images using context relevance information, MANPU 2016, 1st Int. Workshop on coMics ANalysis, Processing and Understanding, Cancun, Mexique, 4 Décembre 2016, doi : 10.1145/3011549.3011561
- > Christophe Rigaud, Srikanta Pal, Jean Christophe Burie, Jean-Marc Ogier, Toward speech text recognition for comic books, MANPU 2016, 1st Int. Workshop on coMics ANalysis, Processing and Understanding, article n°8, p 8:1-8:6, Cancun, Mexique, 4 Décembre 2016, doi : 10.1145/3011549.3011557.
- > Christophe Rigaud, Thanh Nam Le, Jean Christophe Burie, Jean-Marc Ogier, Shoya Ishimaru, Motoi Iwata, Koichi Kise, Semi-automatic Analysis of Manga Using Eye Tracking information, DAS 2016, Twelveth IAPR International Workshop on Document Analysis Systems, Santorin, Grece, 11-14 Avril 2016.
- > Christophe Rigaud, Jean Christophe Burie, Jean-Marc Ogier, Text-independent speech balloon segmentation for comics and manga, GREC 2015, 11th IAPR International Workshop on Graphics Recognition, Nancy, France, 22-23 Août 2015.
- > Thanh Nam Le, Muhammad Muzzamil Luqman, Jean-Christophe Burie, J.-M. Ogier, A Comic Image Retrieval System Using Multi-layer Graph Representation and Frequent Subgraph Mining Techniques, Gbr2015, Workshop on Graph-based Representations in Pattern Recognition, Beijing, Chine, 13-15 Mai 2015
- > Christophe Rigaud, Jean Christophe Burie, Dimosthenis Karatzas, Jean-Marc Ogier, Color descriptor for content-based drawing retrieval, DAS 2014, Eleventh IAPR International Workshop on Document Analysis Systems, Tours, France, 7-10 Avril 2014.
- > Muhammad Muzzamil Luqman, Hoang Nam HO, Jean Christophe Burie, Jean-Marc Ogier, Automatic indexing of comic page images for query by example based focused content retrieval, GREC 2013, 10th IAPR International Workshop on Graphics Recognition, Bethlehem, Etats-Unis, 20-21 Août 2013.
- > Christophe Rigaud, Dimosthenis Karatzas, Jean Christophe Burie, Jean-Marc Ogier, Speech balloon contour classification in comics, GREC 2013, 10th IAPR International Workshop on Graphics Recognition, Bethlehem, Etats-Unis, 20-21 Août 2013.
- > Hoang Nam HO, Christophe Rigaud, Jean Christophe Burie, Jean-Marc Ogier, Redundant structure detection in attributed adjacency graphs for character detection in comics books, GREC 2013, 10th IAPR International Workshop on Graphics Recognition, Bethlehem, Etats-Unis, 20-21 Août 2013.
- > Anh Khoi Ngo Ho, Jean-Christophe Burie, Jean-Marc Ogier, Panel and speech balloon extraction from comic books, DAS 2012, Tenth IAPR International Workshop on Document Analysis Systems, Gold Coast, Australie, 27-29 Mars 2012.

ED I2S : Module image 2019

19 avril 2019 - Jean-Christophe Burie

