


RAPPORT FINAL DU PROJET2

Thème : reconnaissance de scène naturelle

Membres : **BIAKOTA BOMBIA Herbert Cepha**
MILORME Pierre Rubens
PERRAULT André

Introduction

Durant toutes ces années l'informatique a fait de progrès considérable dans le plus domaine scientifique et la technologie. En dépit de toutes ces avancées jusqu'à de nos jours, plusieurs recherche ont été axée dans le domaine de l'intelligence artificiel applique à l'indexation du contenu multimédia et de la numérisation, ainsi, applique ce dernier à la classification de la scène naturelle peut s'avérer beaucoup mieux efficace. Cependant, reconnaissance de scènes au niveau des systèmes artificiels demeure encore très difficile. Ainsi, pour pouvoir bien arriver à bien reconnaître les différents éléments que constituent une image de scène naturelle afin de pouvoir classifier il primordial d'avoir cette problématique qui est de savoir "Comment catégoriser ou classifier une image de scènes naturelles ?". Nous avons réduit la question en "Selon quels critères une image fait partie d'une catégorie, et pas d'une autre ?". Ainsi ces questions vont nous de référentiel pour la réalisation de ce travail

1- État de l'art

Durant cette dernière année beaucoup de progrès ont été fait dans le domaine de la numérique. Le but de ce travail est de faire un résumé exhaustif de l'ensemble des méthodes déjà existantes dans le domaine. Pour ce fait, Il existe de nombreuses façons dont les travaux antérieurs ont abordé ce problème. utilise l'idée traditionnelle de l'approche ascendante où la reconnaissance des scènes se fait par segmentation de l'image dans les régions plutôt que pour les fonctionnalités globales. [10] va pour une approche typique. [8] élimine la nécessité d'annotations humaines des fonctionnalités d'image. [2] introduit le concept de la reconnaissance des scènes sur la correspondance géographique géométrique approximative et prétend obtenir des taux de reconnaissance élevés sur les données difficiles. souligne le fait que nous reconnaissons les images à différentes échelles et pas seulement avec des fonctionnalités globales ou uniquement avec des fonctionnalités locales et fait un très bon travail de classement des images.

En conséquence, la diversité des problèmes évoqués a induit le développement des solutions fondées sur des approches très différentes Plusieurs états de l'art ont été publiés ces dernières années sur ce sujet. Les auteurs de [Radke et al] proposent une description des méthodes générales employées en détection dans une image contenant une scène, mais ne traitent pas des méthodes de « comparaison post classification » (*delta classification*), c'est-a-dire des méthodes consistant à classifier les différentes images pour ensuite détecter des différences sur les images classifiées. Ils s'intéressent au problème de détection de point caractéristiques du point de vue méthodologique, mais ne s'intéressent pas à des

applications particulières. Cette publication propose aussi quelques pistes en ce qui concerne le travail difficile de l'évaluation de performance en détection de caractéristiques. D'autres publications, sont axées sur des problématiques plus particulières comme les détections de changements sur les images aériennes ou la surveillance écosystèmes. Plus récemment, dans le cadre d'une thèse de doctorat, A. Robin présente une taxonomie des différentes méthodes de détection de changements dans [Robin, 2011,]

1-1 Approches (contextuelles)

Les approches fondées sur les pixels ont des limitations importantes dues à leur nature même. Plusieurs auteurs ont proposé d'utiliser des informations contextuelles ou géométriques afin d'améliorer la qualité de la détection. On cherche avec ces approches, non plus à détecter des pixels variables, mais des zones plus étendues. Ceci limite également le nombre de pixels isolés qui pourront faire l'objet d'une mauvaise classification. On peut faire appel à des structures de plus ou moins haut niveau d'interprétation pour améliorer la classification. On peut citer notamment les champs de Markov, les lignes de niveau, la notion de texture, ou des objets de haut niveau d'interprétation.

1-2 Subdivision de l'image

Parmi les méthodes contextuelles classiques, une des plus simples consiste à diviser l'image en zones régulières (le plus souvent à l'aide d'une grille, parfois à l'aide d'une segmentation préalable) et à donner une estimation du changement au niveau de chaque zone et non plus au niveau des pixels. Ce type d'approches présente deux défauts principaux : une perte de précision importante et une segmentation relativement arbitraire (dans le cas de la grille) ou un choix difficile au niveau de la segmentation à adopter dans le second cas.

1-2 Voisinages

On a souvent recours à la notion de voisinage, pour avoir une carte de changements régulière et éviter les phénomènes de bruit. Ainsi, dans , chaque entrée du réseau de neurones [11] n'est pas un pixel isolé, mais un pixel accompagné de son voisinage 8-connexe. Cet exemple est assez révélateur de la façon dont la notion de voisinage est intégrée au sein des systèmes plus complexes de détection de changements.

1-3 Ensemble de niveau

Une autre notion de voisinage, mais basée sur les ensembles de niveau a été utilisée en détection de caractéristiques. Chaque image est transformée en un arbre de formes qui sont les composantes connexes des ensembles de niveau. La détection de changements se fonde sur la reconnaissance de similarités entre les

deux arbres. Cette technique semble donner des résultats convaincants sur des images de vidéo surveillance en milieu urbain. Nous reviendrons un peu plus en détail sur ces techniques

1-4 classification

Certains auteurs ont eu recours aux techniques de comparaison « post classification » on utilise les informations statistiques liées à la texture pour classer, à l'aide d'un réseau de neurones, des parcelles d'images en deux catégories : urbain et rural. Une fois cette étape effectuée sur des couples d'images, les zones correspondantes de chaque paire sont confrontées afin de détecter des changements via des classifications contradictoires. Afin d'éviter des fausses alarmes, seuls des groupes de zones connexes sont finalement retenus. Cette approche présente plusieurs inconvénients : d'une part, elle nécessite un apprentissage du réseau de neurones et deuxièmement, elle ne permet pas une localisation fine des changements intervenus. Dans certains cas, la segmentation *a priori* a un intérêt dans le sens où les parcelles segmentées représentent directement l'information recherchée. De manière plus générale, ces techniques présentent l'inconvénient majeur de requérir un traitement préalable de chaque image du couple ou de la série multi temporelle, et par conséquent la résolution d'un problème non-trivial de détection ou de classification. L'information de différence n'est en général pas utilisée dans ce traitement préalable. En revanche, elles sont largement utilisées en coordination avec un Système d'Information Géographique (SIG), souvent, pour mettre à jour les informations contenues dans les SIG. On fait alors appel assez souvent à des méthodes de haut niveau introduisant des relations entre les objets présents dans le SIG et les objets détectés par la prise de vue. les auteurs décrivent des relations topologiques qualitatives entre les objets obtenus à l'aide, entre autres, du système *région connexion calculs* (RCC-8). Ce système décrit six relations possibles entre deux objets : disjonction, tangence, intersection partielle, égalité stricte, tangence interne et inclusion stricte. Les deux dernières relations sont réciproques, portant le total à huit

1-5 Méthodes d'extraction des caractéristiques

A ce niveau, il est très important de savoir les différentes méthodes pour la réalisation de l'extraction des caractéristiques d'une image de scène naturelle à fin de pouvoir la classer. nous avons: Descripteur SIFT, SCIP (Shape context for Interest Point), USAN etc.

1-6 Méthodes de classification des images

KNN (Les K plus proches voisins), K-means, SVM Les "support Vector machine" Classifieurs bayésiens, le reseau de neurone, decision tree

1-6-1 Réseau neurone [11-13]

Avant de mener une étude multi-simulation globale, le réseau de neurones est d'abord divisé en plusieurs réseaux sous-neuronaux qui se composent de deux couches et forme ces réseaux sous-neurales couche par couche, puis la valeur initiale des réseaux neuronaux multicouches sera obtenue en accumulant celle de ces sous-réseaux neuronaux et enfin en utilisant un algorithme d'optimisation global tel que l'algorithme BP pour bien régler. L'étude de la profondeur en tant que recherche émergente dans le domaine de l'apprentissage par machine a été largement concernée par le milieu universitaire et l'industrie. Il a également obtenu de bons résultats dans la reconnaissance de l'image, la reconnaissance de la parole, le processus du langage naturel, la récupération d'informations et d'autres domaines

1-6-2 SVM Les “support Vector machine”

Un système de classification de la scène basé sur SVM est comparée au système d'analyse discriminante linéaire (LDA) et les résultats favorisent le modèle SVM avec un pourcentage de précision élevé. Dans les expériences relatives, le modèle SVM surpasse constamment les autres modèles en termes d'efficacité et de précision. La sélection des fonctionnalités pour la classification est un domaine de recherche très actif dans l'exploration de données et l'optimisation. Sa nature combinatoire nécessite le développement de techniques spécifiques (comme les filtres, les enveloppes, les algorithmes génétiques, etc.) ou des approches hybrides combinant plusieurs méthodes d'optimisation. Dans ce contexte, l'élimination des fonctionnalités récursives de la machine vectorielle de support (SVM-RFE) se distingue comme l'une des méthodes les plus efficaces. Cependant, l'algorithme RFE-SVM est une méthode gourmande qui espère seulement trouver la meilleure combinaison possible pour la classification. Pour surmonter cette limitation, nous proposons une approche alternative dans le but de combiner l'algorithme RFE-SVM avec les opérateurs locaux de recherche en fonction de la recherche opérationnelle et de l'intelligence artificielle.

II- Solution proposée

1- Démarche

Notre pipeline de reconnaissance de scène implique des étapes d'extraction, de codage, de mise en relation et de classification. Notre modèle a plusieurs paramètres accordables tels que la taille du vocabulaire visuel (K), le noyau à utiliser par le classificateur SVM (par exemple linéaire par rapport à Gaussien) ainsi que les hyper-paramètres SVM tels que C et γ . Nous utilisons une validation croisée pour choisir empiriquement les hyper-paramètres optimaux. Nous

découvrons d'abord le K optimisé pour un SVM linéaire. Nous utilisons ce K et effectuons ensuite une recherche de grille pour découvrir le C et γ optimal correspondant à un modèle SVM ayant un noyau de fonction de base radiale (RBF). Les détails de chacune de nos étapes de pipeline sont présentés ci-dessous.

Notre projet se concentre principalement sur la classification d'une scène naturelle et, en tant que tel, il est impératif que nous clarifions ce que nous entendons par une scène. Pour distinguer la scène de «objet» ou de «texture», nous suivons l'approche adaptée et nous considérons la distance absolue entre l'observateur et la zone fixée comme facteur discriminant. Ainsi, un «objet» est quelque chose qui subisse environ 1 à 2 mètres autour de l'observateur; Mais dans le cas d'une scène, la distance entre l'observateur et le point fixé est généralement supérieure à 5 mètres. Pour le dire en commun, l'objet est quelque chose qui est à portée de main, alors que la scène est principalement un endroit où l'on peut se déplacer.

1.1 Acquisition

Nous avons utilisé la base de données « 13 Natural Scene Categories » pour la mise en œuvre de notre système. Cette base décrit 13 types de scènes naturelles représentées par des images en niveau de gris. Dans notre premier scénario d'expérimentation les 100 premières images de chaque classe servent à l'entraînement et le reste pour le test. Dans le deuxième scénario, nous classons toutes les images en deux grands groupes « intérieur » et « extérieur » puis nous utiliserons la moitié des images de chaque groupe pour l'apprentissage et l'autre moitié pour les tests.

1.2 Extraction de caractéristiques

De chaque image, nous extrayons les caractéristiques de DAISY [3] qui correspondent aux points-clés détectés à partir de l'image. Nous utilisons la bibliothèque skimage (scikit-image) [4] pour l'extraction des fonctionnalités. Correspondant à chaque point clé de l'image, nous extrayons un descripteur DAISY. Nous extrayons également un descripteur HOG standard correspondant à l'image entière à une granularité différente (en utilisant les paramètres pixels par cellule, cellules par bloc et orientations), ce qui nous permet effectivement de choisir des caractéristiques à différentes échelles. Nous utilisons les descripteurs DAISY extraits de toutes les images d'entraînement pour former un bag-of-visual-words en les regroupant à l'aide de Mini-Batch K-Means. Nous déterminons empiriquement la meilleure taille de vocabulaire visuel (K) par validation croisée.

1.3 Codage

Nous utilisons les fonctionnalités DAISY locales correspondant à chaque point clé pour encoder chaque image comme un histogramme de mots visuels. Nous utilisons ici le concept standard de «bag-of-visual-words». Concrètement, nous appliquons un algorithme K-means pour quantifier les fonctionnalités DAISY dans des grappes 'K' pour former des "mots visuels" dans un vocabulaire. K représente la taille du vocabulaire. Correspondant à chaque image, nous formons un histogramme avec 'K' comme dimensionnalité, en utilisant ce vocabulaire. Cette représentation codée de l'image forme l'histogramme "DAISY" et constitue une partie de notre descripteur de caractéristiques hybrides.

1.4 Mise en relation

Nous utilisons un schéma de mise en commun à 2 niveaux dans notre pipeline de reconnaissance de scène. Du point de vue DAISY, nous construisons un histogramme en représentant la fréquence de chaque mot visuel dans chaque image. Nous effectuons cela en sélectionnant chaque point clé de l'image et en recherchant l'identifiant de cluster correspondant à ce descripteur DAISY et en incrémentant le compte correspondant à ce bac dans notre «histogramme DAISY». Cette procédure est effectivement un «pooling de somme». Nous effectuons une normalisation L2 de l'histogramme résultant pour former une fonction d'histogramme DAISY que nous appelons «histogramme DAISY». Pour la mise en commun du 2ème niveau, nous prenons le descripteur global HOG correspondant à chaque image et effectuons une normalisation L2 suivie d'une concaténation avec la caractéristique d'histogramme DAISY correspondante, pour former notre descripteur de caractéristiques hybrides.

1.5 Classification

Nous utilisons le classificateur SVM standard avec divers noyaux, pour la classification. Nous utilisons la bibliothèque sklearn (scikit-learn) [5] pour les implémentations SVM. Nous effectuons une validation croisée en divisant de manière aléatoire l'ensemble de données en un ensemble de formation et de validation. Nous construisons le «vocabulaire visuel» ainsi que les vecteurs de fonctionnalités de formation de la division de formation. Nous rapportons la précision globale, la confusion-matrice ainsi que les statistiques standard de récupération d'informations telles que la précision, le rappel et la mesure F correspondant à nos expériences.

1.6 - Évaluation du système

Pour évaluer notre modèle, nous utiliseront les mesures que sont le taux de bon classement et la matrice de confusion.

III- l'expérimentation

Fonctionnement et exécution du programme

Pour pouvoir utiliser du programme, il est nécessaire d'effectuer les actions à priori telles que :

- installer anaconda 3

install les packages: **sklearn, skimage, tqdm**

Le programme implémenté est nommé **reconnaissance_de_scenes.py**. Ainsi pour l'exécuter nous pouvons le faire de 2 manières:

- il suffit de se positionner dans le répertoire « projet2 » par ligne de commande. Et exécutez la commande suivant : **python3 reconnaissance_de_scenes.py**

- il faut utiliser Jupyter puis se positionner dans le répertoire ou se trouve le fichier et importer le contenant le code source à l'aide de la commande **%load reconnaissance_de_scenes.py** et ensuite taper sur la combinaison des touches **shift et enter**.

Résultat avec jupyter :

pour l'histogramme de gradient

```
Methode de classification avec HOG
Nombre d instances test 2556 2556
Precision globale: 0.616588419405
```


	precision	recall	f1-score	support
kitchen	0.46	0.48	0.47	110
livingroom	0.54	0.44	0.49	189
MITforest	0.59	0.88	0.70	228
bedroom	0.36	0.43	0.40	116
MITopencountry	0.56	0.44	0.50	308
MITmountain	0.59	0.60	0.59	274
MITcoast	0.67	0.82	0.73	260
MITstreet	0.67	0.78	0.72	192
PARoffice	0.60	0.44	0.51	115
MITtallbuilding	0.71	0.59	0.65	256
MITinsidecity	0.62	0.65	0.63	208
MITHighway	0.77	0.38	0.50	160
CALsuburb	0.88	0.91	0.89	140
avg / total	0.62	0.62	0.61	2556

Taux de

précision et rappel pour chaque chaque prédite avec comme support le nombre d'objets qui les contient

Matrice de confusion:

```
[[ 53  24   0  18   0   1   1   2   4   3   4   0   0]
 [ 15  84   1  54   1   1   0   6  12   6   6   1   2]
 [   0   0 201   0   2  24   0   0   0   0   0   1   0]
 [   9  29   3  50   0   3   0   4   4   4   5   2   3]
 [   0   0  41   0 136  62  56   7   1   1   1   3   0]
 [   0   0  64   0  29 165   7   2   1   4   0   2   0]
 [   0   0   5   0  26   5 212   3   2   0   0   5   2]
 [   1   2   7   0   6   5   0 150   2   9   5   3   2]
 [  26   8   0  12   0   1   0   5  51   5   5   0   2]
 [   6   7  10   2   2   6   1   7   6 152  54   1   2]
 [   4   1   2   0   2   1   3  28   1  28 135   0   3]
 [   0   0   1   1  37   7  38  10   1   1   2  60   2]
 [   0   1   8   0   0   0   0   1   0   2   1   0 127]]
```

nous remarquons que pour algorithme d'Histogramme de Gradient la prédiction nous donne un taux de précision de 61.65%

avec DAISY

Methode de classification avec DAISY
 Nombre d instances test: 2556 2556
 Precision globale: 0.711267605634

	precision	recall	f1-score	support
kitchen	0.38	0.36	0.37	110
livingroom	0.48	0.50	0.49	189
MITforest	0.78	0.94	0.85	228
bedroom	0.42	0.43	0.42	116
MITopencountry	0.70	0.64	0.67	308
MITmountain	0.78	0.80	0.79	274
MITcoast	0.73	0.79	0.76	260
MITstreet	0.75	0.87	0.80	192
PARoffice	0.71	0.70	0.70	115
MITtallbuilding	0.87	0.69	0.77	256
MITinsidecity	0.71	0.67	0.69	208
MIThighway	0.80	0.63	0.70	160
CALsuburb	0.89	0.96	0.92	140
avg / total	0.71	0.71	0.71	2556

Taux de précision et rappel pour chaque classe prédite et avec comme support le nombre d'objets qui les contient

Matrice de confusion:

```
[[ 50  23   0  25   0   0   1   1   3   1   5   0   1]
 [ 15  90   0  34   0   1   1   7  18   7  11   2   3]
 [   0   0 219   0   1   7   0   0   0   0   0   0   1]
 [ 12  31   2  55   1   1   1   1   5   1   0   3   3]
 [   0   1  27   0 184  25  52   5   0   3   3   4   4]
 [   0   1  21   1  10 221   4   6   0   7   0   3   0]
 [   0   0   3   1  27   5 215   0   2   0   3   4   0]
 [   0   0   2   1   2   8   0 166   0   6   1   4   2]
 [ 11  10   0  12   0   0   1   0  77   2   1   0   1]
 [   4  11   8   3   0   7   0   5   0 186  23   4   5]
 [   7   2   2   1   3   0   4  34   3   4 143   1   4]
 [   1   0   2   0  29   7  18   6   0   0   2  93   2]
 [   0   3   0   0   0   0   0   0   0   0   0   0 137]]
```

nous remarquons que pour algorithme de DAISY la prédiction nous donne un taux de précision de 71.12%

avec Hybride

Methode de classification avec Hybride

Nombre d instances test 2556 2556

Precision globale: 0.754303599374

	precision	recall	f1-score	support
kitchen	0.55	0.53	0.54	110
livingroom	0.55	0.61	0.58	189
MITforest	0.81	0.90	0.85	228
bedroom	0.47	0.47	0.47	116
MITopencountry	0.66	0.72	0.69	308
MITmountain	0.80	0.78	0.79	274
MITcoast	0.78	0.84	0.81	260
MITstreet	0.83	0.85	0.84	192
PARoffice	0.80	0.75	0.77	115
MITtallbuilding	0.92	0.76	0.83	256
MITinsidecity	0.73	0.75	0.74	208
MIThighway	0.88	0.63	0.73	160
CALsuburb	0.99	0.98	0.98	140
avg / total	0.76	0.75	0.75	2556

Taux de précision et rappel pour chaque classe prédite et avec comme support le nombre d'objets qui les contient

Matrice de confusion:

```
[[ 58  32   0  14   0   0   0   0   2   0   4   0   0]
 [ 12 116   0  36   1   0   0   1  14   0   8   1   0]
 [   0   0 205   0  16   7   0   0   0   0   0   0   0]
 [ 10  41   1  54   1   1   0   1   4   2   0   0   1]
 [   0   1  16   0 223  26  39   1   0   0   0   2   0]
 [   0   1  19   0  27 214   5   2   0   2   1   3   0]
 [   0   0   2   1  36   2 219   0   0   0   0   0   0]
 [   0   1   3   1   3   6   0 164   0   6   5   3   0]
 [ 16   7   0   5   0   0   0   0  86   1   0   0   0]
 [   3   9   3   1   1   5   1   1   1 195  35   1   0]
 [   7   0   1   2   2   0   3  25   1   6 156   4   1]
 [   0   0   2   1  28   8  13   3   0   0   4 101   0]
 [   0   2   0   0   0   0   0   0   0   1   0   0 137]]
```

nous remarquons que pour algorithme de mis en relation (hybride) la prédiction nous donne un taux de précision de 75.43%

Avec Hybride et SVM lineaire

```
Methode de classification avec Hybride - SVM Lineaire
Nombre d instances test 2556 2556
Precision globale: 0.751956181534
```

	precision	recall	f1-score	support
kitchen	0.57	0.48	0.52	110
livingroom	0.59	0.62	0.61	189
MITforest	0.81	0.92	0.86	228
bedroom	0.48	0.46	0.47	116
MITopencountry	0.70	0.65	0.68	308
MITmountain	0.78	0.81	0.79	274
MITcoast	0.72	0.82	0.77	260
MITstreet	0.80	0.89	0.84	192
PARoffice	0.81	0.78	0.80	115
MITtallbuilding	0.90	0.77	0.83	256
MITinsidecity	0.70	0.78	0.74	208
MIThighway	0.89	0.61	0.72	160
CALsuburb	0.94	0.99	0.96	140
avg / total	0.75	0.75	0.75	2556

Taux de précision et rappel pour chaque classe prédite et avec comme support le nombre d'objets qui les contient

Matrice de confusion:

```
[ [ 53 32 0 17 0 0 1 0 1 0 6 0 0]
  [ 12 118 0 30 0 1 1 3 11 3 7 1 2]
  [ 0 0 209 0 9 10 0 0 0 0 0 0 0]
  [ 11 36 2 53 0 1 1 1 4 2 3 1 1]
  [ 0 1 18 0 201 26 51 5 0 0 2 3 1]
  [ 0 0 19 1 13 221 9 4 0 5 1 1 0]
  [ 0 0 2 1 36 3 213 0 3 0 0 2 0]
  [ 0 0 1 0 0 6 0 170 0 4 7 3 1]
  [ 13 2 0 8 0 0 1 0 90 1 0 0 0]
  [ 0 8 4 0 2 5 0 2 0 197 37 0 1]
  [ 4 0 1 1 2 1 2 23 2 7 162 1 2]
  [ 0 2 2 0 23 10 15 5 0 0 5 97 1]
  [ 0 1 0 0 0 0 0 0 0 0 1 0 138]]
```

nous remarquons que pour algorithme de mis en relation (hybride + SVM) la prédiction nous donne un taux de précision de 75.19% et que les nombres d'objets mal prédits pour chaque classe est plus petit

	MITinsidecity	bedroom	PARoffice	MITmountain	MITtallbuilding	MIThighway	MITcoast	livingroom	MITopencountry	MITstreet	MITforest	kitchen	CALsuburb
MITinsidecity	53	32	0	17	0	0	1	0	1	0	6	0	0
bedroom	12	118	0	30	0	1	1	3	11	3	7	1	2
PARoffice	0	0	209	0	9	10	0	0	0	0	0	0	0
MITmountain	11	36	2	53	0	1	1	1	4	2	3	1	1
MITtallbuilding	0	1	18	0	201	26	51	5	0	0	2	3	1
MIThighway	0	0	19	1	13	221	9	4	0	5	1	1	0
MITcoast	0	0	2	1	36	3	213	0	3	0	0	2	0
livingroom	0	0	1	0	0	6	0	170	0	4	7	3	1
MITopencountry	13	2	0	8	0	0	1	0	90	1	0	0	0
MITstreet	0	8	4	0	2	5	0	2	0	197	37	0	1
MITforest	4	0	1	1	2	1	2	23	2	7	162	1	2
kitchen	0	2	2	0	23	10	15	5	0	0	5	97	1
CALsuburb	0	1	0	0	0	0	0	0	0	0	1	0	138

Pour ce fait si nous nous basons sur la matrice de confusion pour faire le comparaison des différentes expérimentations (HOG, DAISY, HYBRIDE(hog et DAISY), HYBRIDE et SVM) que nous avons implémentés:

- au niveau des classes mâles du point de vue visuel pour l'expérimentation avec hybride et SVM, nous avons moins d'objets mal prédits pour chaque classe ce qui donne plus d'objets bien prédits, ce que s'explique aussi par un bon taux de classement de chaque classe et un bon taux de classement global

Conclusion

pour notre travail, nous avons pu ressortie quelques solutions existantes pour ce sujet, cependant nous avons défini un pipeline réalisation de notre travail basée sur des méthodes qui prend en compte tous les objets ainsi que le fond d'une image (l'image d'une scène naturelle)

Références:

- [2] Oliva A. and Schyns PG, “Diagnostic colors mediate scene recognition,” *Cognitive Psychology*, vol.41, pp. 176–210, 2011.
- [8] Le Saux B. and Amato G., “Image classifier for scene analysis,” *Computer Vision and Graphics International Conference, ICCVG 2004*, Warsaw, Poland, September 2012.
- [10] Julia Vogel , Bernt Schiele , 2013, 'A semantic typicality measure for natural scene categorization', *Pattern Recognition Symposium, DAGM*
11. Hu, X., Zhu, J.: Deep learning: new hotspot in machine learning. *Commun. CCF* 9(7), 64–69 (2013). (in Chinese)
12. Yu, K., Jia, Y., Chen, Y., et al.: Deep learning: yesterday, today, and tomorrow. *J. Comput. Res. Dev.* 50 (9), 1799–1804 (2013). (in Chinese)
13. Krizhevski, A., Sutskever, I., Hinton, G.: Image net classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, vol. 25, pp. 1106–1114 (2012)
- SVM :Yahya Slimani, Mohamed Amir Essegi, Mouhamadou Lamine Samb, Fodé Camara, Samba Ndiaye. Approche de sélection d'attributs pour la classification basée sur l'algorithme RFE-SVM *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, INRIA, 2014, 17, pp.197-219