Reconnaissance d'organismes marins dans des images photographiques sous-marines

Étude bibliographique

Master *Sciences et Technologies*,
Mention *Informatique*,
Parcours DECOL

Auteur
Sébastien Villon

Superviseurs
Marc Chaumont
Jérôme Pasquet
Gerard Subsol
Thomas Claverie
David Mouillot
Sébastien Villéger

Lieu de stage
LIRMM UM5506 - CNRS, Université de Montpellier

8 mars 2015

Résumé

Cette étude présente les différents aspects permettant la détection, la reconnaissance et la classification d'espèces de poissons pour la surveillance des écosystèmes sous-marins. Nous détaillons en particuliers les aspects d'extraction d'images, de suivi, mais aussi les différentes manières de représenter les données extraites. L'article présente aussi une revue des principales méthodes de classification et de reconnaissance, comme les calculs de distances euclidiennes, les classificateurs SVM, les ensembles classifieurs et les réseaux neuronaux, puis met en avant les particularités de ces méthodes sur le thème de la classification d'espèces sous-marines

Abstract

This paper presents the different ways of detecting, recognizing and classifying fish species for marine ecosystems survey. We will see in particular the aspects of extraction, tracking, and many ways of representing the information from extracted datas. This paper also shows the main methods of classification and recognition, like Euclidian distance computation, SVM classifiers, Ensemble classifier, and neural network, then points out the particularities of each of these methods on the submarine species classification theme.

Table des matières

1

Introduction

1.1 Présentation générale

La quantification de l'impact des activités humaines sur l'environnement est un enjeu majeur pour la conservation de la biodiversité des écosystèmes sous-marins. La plus grande difficulté, pour la surveillance de ces écosystèmes, est la constitution de bases de données fiables à grande échelle.
Les techniques de surveillance par extraction comme la pêche ne fournissent que des informations limitées à certaines espèces, et l'interprétation des données récupérées peut être biaisée par l'échantillonnage [19]. De plus, le recours à la pêche, même pour la surveillance, impacte la biodiversité étudiée. Il existe aujourd'hui de nombreuses façons de récupérer les informations sous-marines, nous pouvons les diviser en deux catégories.
Les études "manuelles", nécessitant deux plongeurs prenant des annotations à la main, effectuées lors de la plongée. Cette catégorie est la plus coûteuse en temps et en moyens, et dépend de nombreux facteurs (expérience du plongeur, poissons craintifs...).
Les méthodes reposant sur l'acquisition de vidéos et le comptage de poissons à partir de celles-ci peuvent être effectuées de nombreuses manières (avec ou sans appâts, caméras mobiles ou immobiles). Les techniques les plus utilisées dans le cadre du comptage et de l'étude des espèces par vidéos sont les méthodes RUV (Remote Underwater Vidéo) et DOV (Diver Operated Vidéo)[19]. Les études vidéo enregistrent des quantités de données beaucoup plus importantes (112 teraoctets acquis dans le cadre du projet européen Fish4knowledge [6]) mais sont extrêmement longues à analyser manuellement. C'est dans le but de traiter efficacement ces nombreux enregistrements vidéo que sont envisagées les différentes techniques de traitements automatiques.

La plupart de ces traitements se décomposent en trois phases [23] : la détection, le suivi et la reconnaissance. Notons que certains auteurs [22, 20] regroupent la phase de détection et de reconnaissance en cherchant à détecter directement les espèces de poissons, sans situer au préalable une zone précise de l'image à étudier.

Cette étude bibliographique se divise en quatre parties. Dans un premier temps, nous présentons différentes méthodes utilisées pour la détection et le suivi d'objets. Dans la seconde partie, nous faisons un état de l'art sur les techniques de classification utilisées dans le domaine de la classification d'espèces de poissons (calculs de distance, utilisation d'arbres de décisions, utilisation de classifieurs SVM ou de réseaux de neurones). Dans la troisième partie, nous parlons des méthodes générales d'apprentissage et de classification

FIGURE 1.1 : Déroulement de l'analyse de frames vidéos

peu representées sur la thématique de classification d'espèces sous-marines, comme les méthodes d'apprentissage par *Deep Learning*. Pour finir, nous présentons les différentes contributions que nous comptons apporter au cours du stage.

1.2 Contexte du stage

Notre stage sera effectué en association avec le laboratoire MARBEC (MARine Biodiversity Exploitation & Conservation [1]), spécialisé dans l'étude des changements de la biodiversité dans les écosystèmes marins pour optimiser leur exploitation et leur gouvernance. Dans le cadre de notre étude et de cette collaboration, nous avons décidé de participer au projet international lifeCLEF [2], et plus particulièrement à la tâche *Image Based Fish Identification and Species Recognition* pour laquelle l'organisation met à disposition 285 vidéos labelisées contenant 19868 poissons annotés. Cette compétition correspond parfaitement au projet du stage, à savoir la détection et la reconnaissance d'espèces sous-marines, et nous permettra de comparer notre approche à celles des autres participants, avec une base de connaissance commune. Les données mises à disposition sont un ensemble d'imagettes servant pour l'apprentissage, et 20 vidéos accompagnées d'annotations sous formes d'un formulaire XML, contenant les poissons à identifier. C'est sur cette base que nous appliquerons notre méthode de reconnaissance (Fig 1.1)

Pour notre étude, les difficultés majeures lors de la détection sont les changements de couleur, les différences de lumière, les sédiments présents dans l'eau et les plantes sous-marines (arrière-plan changeant), ainsi qu'une faible qualité des images due à la résolution de base et à la compression [2]. Pour la classification des poissons, des individus d'une même espèce peuvent présenter des différences de taille, d'orientation, de mouvement (positions) en plus des particularités individuelles. Certaines approches décident de ne pas traiter ces problèmes en traitant des images prises dans des conditions contrôlées et standardisées [23], en bassin, de profil, etc.

[1] http ://www.ecosym.univ-montp2.fr/
[2] http ://www.imageclef.org/lifeclef/2015

2

Détection d'une zone d'intérêt et représentation du contenu

2.1 Généralités

La détection de zones d'intérêts, c'est à dire la définition d'une position et d'une zone dans l'image est un pré-traitement qui permet, pendant l'opération de reconnaissance, de ne traiter qu'une partie de l'image susceptible de contenir l'objet à traiter afin de gagner en temps de calcul [7]. Avant de détecter cette zone d'intérêt, il est nécessaire de représenter le contenu de l'image (voir détail partie 2.4).
On peut voir deux catégories de représentations du contenu d'une image, la représentation *pixel* et la représentation *objet*. Pour la représentation orientée pixel [1], chaque pixel est représenté par un vecteur puis associé à un label (poisson ou non-poisson), ce qui permettra, pendant la phase de reconnaissance (définie en partie 3), d'évaluer la probabilité de chaque pixel d'appartenir à l'objet.

Pour la détection de poissons, lors de représentations orientées objets [30], chaque image exemple de l'objet recherché est représentée par un vecteur caractéristique qui servira à l'apprentissage d'un modèle [21]. La représentation orientée objet présente la particularité de pouvoir présenter des caractéristiques de plus haut niveau, permettant de spécialiser les vecteurs caractéristiques.
Dans cette approche, on dispose d'un ensemble N d'imagettes représentant les espèces de poissons. Après avoir été normalisées, ces imagettes peuvent ensuite être décrites grâce à des vecteurs caractéristiques (voir partie 2.4 : Représentation), c'est-à-dire un vecteur représentant les différentes caractéristiques choisies. Le pré-traitement est la plupart du temps [23, 2, 24, 21, 5] composé des même phases, à savoir le passage de l'image en niveau de gris et la normalisation des tailles des objets.

2.2 Détection et suivi d'objet

Le suivi d'objet (tracking) permet un gain de temps au niveau de l'analyse. La plupart des méthodes passent par la recherche d'un motif (boîte englobante) d'une image à une autre [27]. Après une phase d'initialisation, c'est-à-dire la définition de la boîte englobante dans la première image de la séquence, on calcule une représentation de l'objet (abordée en section 2.4) ou on détecte directement les bordures de l'objet (avec un filtre

de Canny par exemple [24]), puis on cherche à les retrouver dans les images suivantes afin d'inférer la trajectoire des objets. Concetto Spampinato [28] propose de baser la représentation sur des matrices de covariance. Pour chaque objet détecté, on calcule une matrice de covariance en créant un vecteur par pixel, composée des coordonnées de celui-ci, de ses valeurs de teintes et de RGB, ainsi que de la moyenne et de la déviation standard d'une fenêtre de 5*5 avec le pixel choisi au centre de la fenêtre. Cette matrice sert ensuite comme modèle de l'objet, puis elle est comparée de frame en frame aux autres objets traqués [22] grâce à un calcul de distance de Forstner [2] qui permet de connaître la similarité entre deux matrices de covariance. Le suivi permet principalement de collecter toutes les images possibles et de les associer directement à la bonne espèce. De nombreuses imagettes d'un même poisson ainsi récupérées peuvent rendre plus robuste la phase de classification (détermination de l'espèce). Le suivi permet également dans certains cas de mieux gérer les occultations et donc d'obtenir de meilleurs résultats qu'une simple détection [7]. Il est à noter que la gestion d'apparition et de disparition de poissons dans la séquence d'images est difficile à gérer, mais ce problème est en dehors du cadre du stage.

2.3 Initialisation du suivi et de la détection

L'approche vidéo mélange les phases de suivi et de détection. Cette approche utilise une méthode de soustraction de l'arrière plan [20, 24, 27, 5] des frames, afin de localiser les objets mouvants. Pour détecter ce qu'est l'arrière plan, et en connaissant les problèmes liés au milieu traité, plusieurs axes sont envisagés : la détection par modèles gaussiens [20, 27], en utilisant l'algorithme GMM (Gaussian Mixture Models), ou la création d'un modèle de l'arrière plan estimé à partir de la modification des pixels d'une frame à l'autre [5].

Une des idées principales du GMM pour éviter de prendre un déplacement d'objets de l'arrière plan pour un poisson à identifier est la mise en place d'un "palier de déplacement"[5]. En effet, en considérant chaque pixel modelé par un mélange de distributions gaussiennes l'article définit un pixel comme appartenant à l'avant-plan si sa déviation est supérieure de 2.5 à la moyenne de toutes les autres distributions. À la fin des opérations de suppressions d'arrière plan, on peut délimiter une zone de l'image particulièrement susceptible de contenir un objet que l'on cherche à classer.

Une autre méthode pouvant venir en complément de la soustraction d'arrière plan (ou en pré-traitement) consiste à utiliser une approche d'initialisation de la détection par apprentissage basée sur une méthode de fenêtre glissante. Cette méthode est composée de deux phases. La première phase est l'apprentissage de modèles (histogrammes de teinte, de gradients[14]) représentants l'objet que l'on cherchera à identifier. La seconde phase consiste à déplacer une fenêtre de taille variable à l'intérieur de l'image dans laquelle on cherche à détecter notre objet. Chaque déplacement de la fenêtre glissante crée une sous-image, qui sera classifiée indépendamment. Le but de cette méthode est de faire correspondre une des sous-images avec un modèle d'objet, afin de délimiter l'espace de recherche et d'initialiser la première fenêtre englobante, dans le cadre du suivi.

2.4 Représentation

On dénombre plusieurs représentations *objets* redondantes dans la littérature de classification d'espèces de poissons. Certaines approches se concentrent sur les propriétés de textures [24, 6] en s'appuyant sur des matrices de convolutions, différents types de filtres (half-wave, deuxième convolution), puis concatènent les résultats obtenus dans un vecteur. D'autres approches se concentrent sur la distribution des teintes, les contours, sur des caractéristiques de textures ou sur la représentation *Scale-invariant feature transform* (SIFT). Les vecteurs SIFT déterminent sur les images des zones centrées autour de *points-clés* [5], puis les représentent grâce à un histogramme des orientations locales. Ces descripteurs présentent l'avantage d'être invariants à l'orientation et à la résolution de l'image [17]. Les approches SIFT sont particulièrement robustes aux variations d'échelles, de rotations et d'illuminations, qui sont les problèmes principaux présents lors de la détection en millieu sous-marin.
Quelques cas particuliers [23] utilisent des caractéristiques de ratios par rapport à une référence, mais les conditions dans lesquelles ces études ont été faites (poisson posé sur un fond blanc à une distance connue) et ne correspondent pas à une étude en conditions réelles. Un autre type d'approche [21] représente l'objet sous forme d'un histogramme de gradients orientés (HOG).

Le principe de l'histogramme de gradient orienté est de décrire un objet dans une image grâce à ses contours et aux formes qui le composent en utilisant la distribution d'orientation et d'intensité du gradient. L'article [21] montre que l'utilisation de vecteurs caractéristiques HOG permet de meilleurs résultats dans le cadre de la détection dans des situations peu favorables (objets incomplets, difficiles à reconnaître), ce qui correspond aux difficultés rencontrées lors de la détection et de la reconnaissance en milieu sous-marin : poissons cachés ou partiellement présents, bruit important sur l'image.
La plupart des algorithmes de détection utilisent des vecteurs représentatifs particuliers (tailles, ratios de dimensions, teintes), ou des représentations SIFT [20, 27, 5] ou *shape context* (SC [24]). La représentation de données par des vecteurs HOG est plus efficace [10] que les représentation SIFT et SC [31] dans le cadre de la détection de personnes. Pour ce stage, nous avons choisi d'utiliser une représentation HOG.

Méthodologie d'extraction des HOG

Pour extraire l'histogramme de gradient orienté d'une région, on la divise tout d'abord en cellules de tailles réduites. Pour chaque pixel de cette cellule, on calcule l'orientation de son gradient, puis on crée un histogramme des orientations du gradient pour lequel chaque pixel vote pour une orientation principale. Le vote de chaque pixel est pondéré par son intensité, ce qui permet de mettre en avant les contours les plus marqués dans l'image. La combinaison des histogrammes des cellules forme alors le descripteur HOG d'une région.

3

Reconnaissance

3.1 Introduction

Il existe de nombreuses méthodes pour classer les modèles obtenus. Dans cette partie, nous expliquons quelques opérateurs simples, des classifieurs plus complexes, ainsi que les résultats obtenus pour les deux catégories.

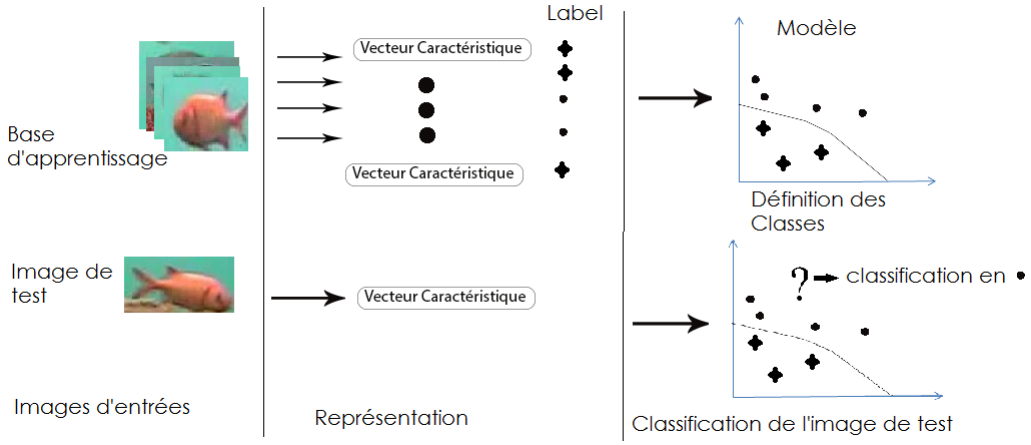
3.2 Approche naïve

Un comparateur courant est le calcul de distance euclidienne [23]. Il s'agit de calculer cette distance entre deux vecteurs représentatifs celui d'un modèle et celui d'une image traitée. On suppose que le modèle connu qui permet d'obtenir la plus petite distance par rapport à l'objet étudié est celui qui lui ressemble le plus. Cette méthode montre de bons résultats lorsque les objets comparés sont très reconnaissables et distincts, mais elle est généralement dépassée par les méthodes d'apprentissages [23].

3.3 Classification

Lors d'une classification, la première phase est la phase d'apprentissage, durant laquelle l'utilisateur choisit les exemples à fournir à l'algorithme d'apprentissage. À l'aide d'une vérité terrain on communique des exemples positifs représentant le poisson sous différents angles et dans différentes positions [24], mais aussi des exemples négatifs [5] comme des échantillons de l'arrière plan. Une fois la base d'apprentissage construite, nous extrayons les vecteurs caractéristiques par imagettes puis utilisons le classifieur pour trouver un modèle. Ce processus est résumé en figure 3.1.

FIGURE 3.1 : Exemple de classification



Ensemble Classifier

Les *ensemble classifier* (EC) sont des regroupements de classifieurs à faible complexité (weak classifier, WC) dont les résultats se fusionnent (strong classifier). Un des avantages de l'utilisation d'un ensemble classifieur est de rendre les résultats moins dépendants d'une seule fonction de classification. De plus, l'EC est aussi performant que le SVM (voir point suivant), et robuste aux problèmes de grandes dimensions.
Afin de renforcer la prédiction, une stratégie possible consiste à pondérer chaque WC en fonction de son taux d'erreur.

Deux méthodes de reconnaissances utilisant des EC pour la reconnaissance d'objets sont étudiées et comparées dans [20]. La première est adaptée d'une proposition Turk & Pentland [29] qui ont développé un algorithme de reconnaissance faciale basé sur l'analyse de composantes principales (PCA). La seconde est basée sur le logiciel VLFeat, qui permet d'utiliser des procédures Scale Invariant Feature Transform (SIFT) [18]. Les deux EC présentent de bons résultats lorsque les images sont pré-traitées (mise à niveau des intensités, haute qualité des images, etc) et fournissent de meilleurs résultats en utilisant une classification binaire plutôt qu'une classification multiclasse (respectivement 100% d'identification correct avec deux classes contre 40% avec quatre classes)

Support Vector Machine

Concepts généraux

Les machines à vecteurs de support (SVM) [12, 9] font partie des classificateurs les plus utilisés en raison de leur robustesse aux problèmes de grandes dimensions et de leurs bons résultats. Le principe de fonctionnement des SVM est de définir un espace de représentation d'un ensemble de tests puis de chercher à tracer un hyperplan séparateur entre les différentes classes. Pour choisir un hyperplan parmi l'infinité des possibles, on sélectionne celui qui maximise la marge, c'est-à-dire la distance entre l'hyperplan et les échantillons les plus proches.

S'il n'existe pas d'hyperplan dans l'espace de dimension de départ, on cherche à reconsidérer le problème dans un espace de dimension supérieure en appliquant aux vecteurs d'entrée X une transformation non-linéaire ϕ. L'espace d'arrivée $\phi(X)$ est appelé espace de redescription. Formaliser le problème de classification par une fonction ϕ permet de raisonner dans une dimension plus grande et donc de chercher une frontière (hyperplan) non-linéaire.

Protocole expérimental pour l'apprentissage

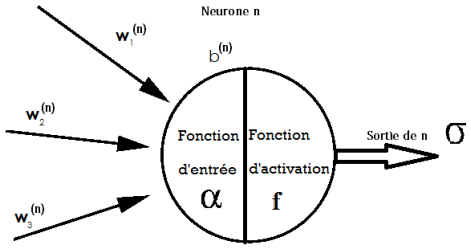
Les articles [24, 5, 22] utilisent des approches de classifications par SVM, soit en utilisant SVMLight [26], soit un SVM classique.
L'article [5] décrit une méthode utilisée lors de la compétition lifeCLEF2014 et cherche à reconnaître 11 espèces de poissons dans 285 vidéos comptant au total 19868 individus. Cet article présente une précision moyenne de reconnaissance de 0.55 et un rappel moyen de 0.35. L'article [24] utilise 320 images tirées de vidéos, soit 160 par espèce, et présente des résultats allant jusqu'à 90% d'objets correctement identifiés. Cette différence de résultats s'explique facilement aux vues des quantités de données traitées ainsi que de la qualité de vidéos utilisées.

Réseaux Neuronaux

Généralités

Les réseaux de neurones artificiels (Artificial Neural Networks, ANN) sont des modèles mathématiques qui tentent de reproduire le comportement du cerveau humain [3]. Un ANN est composé d'un ensemble interconnecté de nœuds, appelés neurones. Chaque neurone possède une fonction d'entrée, une fonction d'activation qui définit l'information qu'il transfèrera aux neurones suivants, une pondération **w** sur chacune des entrées qu'il reçoit, et un biais qui servira de seuil à la fonction d'entrée.

FIGURE 3.2 : Représentation détaillée d'un neurone



Soit α la fonction d'entrée du neurone n, C le nombre de connexions en entrée, $\mathbf{x}$ le vecteur et b le biais du neurone n :

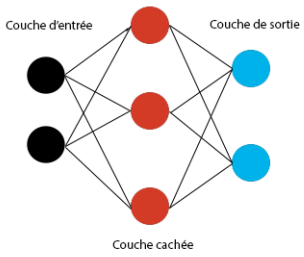
$$\alpha^{(n)}(\mathbf{x}^{(n)}) = \sum_{i=1}^{C} \mathbf{w}_i^{(n)} \mathbf{x}_i^{(n)} + b^{(n)}$$

Soit σ la sortie du neurone n et f sa fonction d'activation :

$$\sigma^{(n)}(\mathbf{x}^{(n)}) = f^{(n)}(\alpha^{(n)}(x)^{(n)})$$

Ces neurones forment des couches. La première couche du réseau est la couche d'entrée, qui reçoit les caractéristiques. La dernière couche est celle de sortie qui correspond aux classes retournées par le réseau. Les couches intermédiaires ou couches cachées contiennent des neurones connectés à tous les neurones de la couche précédente et de la couche suivante (exemple de réseau à une couche en figure 3.3).

FIGURE 3.3 : Réseau de neurones



Pendant l'entraînement d'un ANN, on utilise un mécanisme de rétro-propagation afin de corriger et d'améliorer le réseau, c'est-à-dire modifier les poids associés aux neurones afin que le réseau retourne le résultat attendu :
Soit un vecteur représentant un objet de classe K qui est passé en entrée du réseau, on compare la valeur attendue (100% de probabilité d'appartenir à la classe K) à la valeur obtenue, puis on calcule l'erreur de la couche de sortie. On propage ensuite l'erreur des neurones des couches j aux neurones des couches $j-1$. On met ensuite à jour les poids de chacun des neurones, afin de faire converger le résultat de l'analyse du vecteur de classe K vers un meilleur résultat.

Utilisation

L'utilisation des réseaux neuronaux a montré de meilleurs résultats que les calculs de distances euclidiennes (EDM), en passant de 81% de résultats positifs à 99% dans des conditions identiques [23]. Bien que l'ANN ait des temps de calcul bien supérieurs à l'EDM, de l'ordre de 6.3, la précision du classifieur neuronal est largement supérieure. Les conditions d'analyse de l'article sont particulières puisqu'elles ne sont pas en conditions réelles mais en conditions maitrisées (poisson de côté, à plat, sur fond blanc), toutefois les résultats sont encourageants pour l'utilisation de ce type de classifieurs dans la classification d'espèces sous-marines.

4

Deep Learning

4.1 Introduction

Comme nous l'avons vu précédemment, les ANN sont des ensembles de nœuds appelés neurones. On différencie les réseaux *shallow* (ANN) qui contiennent une seule couche cachée des réseaux *Deep* (DNN) qui en contiennent plusieurs [25].

4.2 Réseaux particuliers et apports des DNN et des CNN

Tout comme les ANN, les DNN utilisent des fonctions de rétro-propagation pour se réajuster pendant l'apprentissage afin de converger. L'ajout de couches cachées d'abstractions soumet les DNN à des erreurs de surentraînement. Pour lutter contre ce surentraînement, ou surapprentissage, les DNN utilisent aussi les fonctions de régularisation comme les fonctions *weight decay* [4] et *dropout* [13].
Les réseaux neuronaux de convolution (CNN) sont utilisés pour permettre une abstraction supérieure. Pour ce faire, une ou plusieurs couches de *convolution* sont utilisées avant les couches *apprentissage* [16]. Chacune des couches de convolution effectue sur ses données d'entrée une transformation grâce à un noyaux de convolution, une étape de non linéarité , puis une étape de *pooling*, qui rend l'apprentissage robuste aux translations. Une fois les n étapes de convolution effectuées, les sorties sont données à un réseau de neurones interconnectés classique (voir exemple figure 4.1).

4.3 Intérêt de l'utilisation du DNN

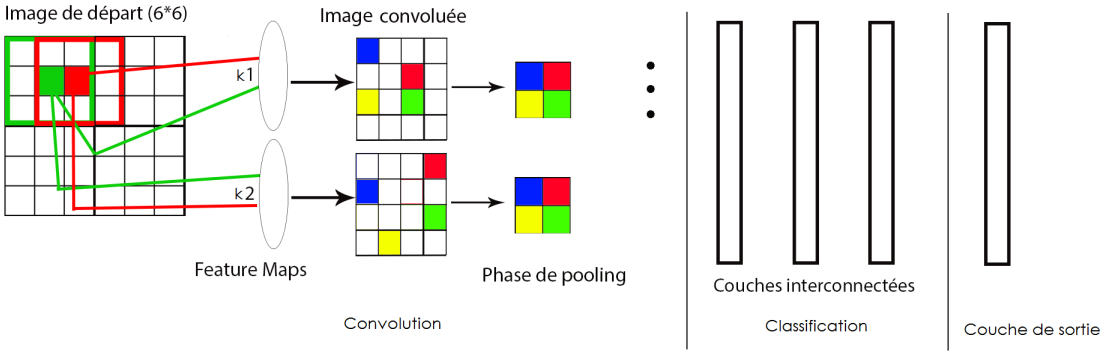
Plusieurs articles utilisent des SVM simples [22, 5, 24]ou des réseaux neuronaux superficiels (shallow-ANN) [23, 2]. Il pourrait être intéressant d'intégrer la technologie des DNN qui semble peu utilisée dans ce contexte afin d'en tester les résultats. Les ANN ayant prouvé leur efficacité [23] sur d'autres types de classification, l'amélioration de ces réseaux grâce aux fonctionnalités des DNN et des CNN semble être encourageante. De plus, les DNN et plus particulièrement les CNN, grâce à leur robustesse et leur niveau d'abstraction, sont très efficaces pour l'analyse et la classification d'image [8], comme cela a été démontré sur la classification d'ImageNet [11]. Les CNN ont aussi fait leur preuve pour des tâches de reconnaissance faciale [15]. La détection d'espèces sous-marines s'inspirant largement des techniques et des méthodes utilisées pour la reconnaissance faciale (

eigenfaces, fisherfaces, représentations PCA et SIFT), il paraît pertinent d'essayer d'appliquer l'apprentissage par CNN pour le challenge d'identification d'espèces de poissons.

Les couches de convolutions fonctionnent ainsi (fig : 4.1)

Dans l'exemple, nous avons une image de 6*6. Pendant la première phase, pour chaque pixel, un neurone n appartenant à une $feature\ map\ f$ applique un kernel de convolution de poids k. Tous les neurones d'une $feature\ map$ partagent le même poids k. À la fin de cette première phase, nous obtenons une image convoluée de même taille, moins les bordures (4*4), pour chaque feature map. Ensuite, toujours pour chaque feature map, une phase de *pooling* est appliquée, dans l'exemple nous appliquons une matrice $MaxPooling$ de 2*2, c'est-à-dire que pour chaque région de 2*2 (sans chevauchement), la valeur de poids maximal est conservée. Après cette étape nous obtenons donc une nouvelle matrice de valeurs représentatives. En fonction de la taille de l'image et des matrices de convolution et de pooling, nous pouvons appliquer plusieurs étapes de convolutions.
À la fin des étapes de convolutions, nous obtenons autant de représentation de l'image de départs qu'il y a de $feature\ maps$, toutes les sorties des $feature\ maps$ sont interconnectées aux couches standards de classification.

FIGURE 4.1 : Réseau de convolution



5

Perspectives et contributions

Lors de ce stage nous tenterons de mettre en application l'utilisation de vecteurs HOG pour améliorer les représentations caractéristiques des objets à identifier, ainsi qu'une méthode d'apprentissage par deep learning.
Nous envisageons plusieurs apports à l'état de l'art dans ce stage.

Pour répondre au problème de traitement automatique de classification d'espèces de poissons, nous proposons de créer et d'utiliser un réseau de type DNN pour la reconnaissance d'espèces sous-marines. Comme nous l'avons vu, l'état de l'art ne propose pas de solutions basées sur le *deep learning* pour répondre au problème de détection et de reconnaissance en milieu sous-marin. Nous proposons donc de mettre en place un CNN, ainsi que d'étudier les différents filtres que nous pouvons appliquer dans le but d'améliorer la classification par ce type de réseau.

L'état de l'art présente peu de classifieurs fusionnant efficacement les différentes caractéristiques possibles à étudier. Nous développerons et intégrerons des caractéristiques spécialisées pour le domaine étudié. En plus de l'utilisation du HOG pour représenter le poisson dans son ensemble, nous pensons que l'ajout de certaines informations telles que la couleur et la répartition des couleurs permettraient d'améliorer la classification.

Nous avons vu qu'un problème majeur des DNN est le temps de calcul. Nous proposons ici d'améliorer le temps de calcul du réseau grâce à une classification intelligente, qui permettra d'éluder des opérations de classification non nécessaire.

Bibliographie

[1] David Aldavert, Arnau Ramisa, Ricardo Toledo, and Ramon López de Mántaras. Efficient Object Pixel-Level Categorization Using Bag of Features. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Yoshinori Kuno, Junxian Wang, Jun-Xuan Wang, Junxian Wang, Renato Pajarola, Peter Lindstrom, André Hinkenjann, Miguel L. Encarnaç ao, Cláudio T. Silva, and Daniel Coming, editors, *Advances in Visual Computing*, volume 5875 of *Lecture Notes in Computer Science*, pages 44–54. Springer, 2009.

[2] Mutasem Khalil Sari Alsmadi, Khairuddin Bin Omar, Shahrul Azman Noah, and Ibrahim Almarashdah. Fish recognition based on the combination between robust feature selection, image segmentation and geometrical parameter techniques using artificial neural network and decision tree. *arXiv preprint arXiv :0912.0986*, 2009.

[3] Peter M Atkinson and ARL Tatnall. Introduction neural networks in remote sensing. *International Journal of remote sensing*, 18(4) :699–709, 1997.

[4] Yoshua Bengio, Nicolas Boulanger-Lewandowski, and Razvan Pascanu. Advances in optimizing recurrent networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8624–8628. IEEE, 2013.

[5] Katy Blanc, Diane Lingrand, and Frédéric Precioso. Fish species recognition from video using svm classifier. In *Working Notes of CLEF 2014 Conference*, 2014.

[6] Bastiaan J Boom, Phoenix X Huang, Cigdem Beyan, Concetto Spampinato, Simone Palazzo, Jiyin He, Emmanuelle Beauxis-Aussalet, Sun-In Lin, Hsiu-Mei Chou, Gayathri Nadarajan, et al. Long-term underwater camera surveillance for monitoring and analysis of fish populations. *VAIB12*, 2012.

[7] Marc Chaumont and William Puech. 3d-face model tracking based on a multi-resolution active search. In *Electronic Imaging 2007*, pages 65081U–65081U. International Society for Optics and Photonics, 2007.

[8] Dan Ciresan, Ueli Meier, and Jürgen Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.

[10] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet : A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[12] Yoav Freund and Robert E Schapire. Large margin classification using the perceptron algorithm. *Machine learning*, 37(3) :277–296, 1999.

[13] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv :1207.0580*, 2012.

[14] Christoph H Lampert, Matthew B Blaschko, and Thomas Hofmann. Beyond sliding windows : Object localization by efficient subwindow search. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[15] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition : A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, 8(1) :98–113, 1997.

[16] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.

[17] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[18] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004.

[19] Delphine Mallet and Dominique Pelletier. Underwater video techniques for observing coastal marine biodiversity : A review of sixty years of publications (1952–2012). *Fisheries Research*, 154 :44–62, 2014.

[20] J Matai, R Kastner, GR Cutter Jr, and DA Demer. Automated techniques for detection and recognition of fishes using computer vision algorithms. In *NOAA Technical Memorandum NMFS-F/SPO-121, Report of the National Marine Fisheries Service Automated Image Processing Workshop, Williams K., Rooper C., Harms J., Eds., Seattle, Washington (September 4–7 2010)*, 2012.

[21] Jerôme Pasquet, Marc Chaumont, and Gérard Subsol. Comparaison de la segmentation pixel et segmentation objet pour la détection d'objets multiples et variables dans les images. *Colloque compression et représentation des signaux Audiovisuels*, 2014.

[22] Fatih Porikli, Oncel Tuzel, and Peter Meer. Covariance tracking using model update based on lie algebra. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 728–735. IEEE, 2006.

[23] C. Pornpanomchai, B Lurstwut, P. Leerasakultham, and W Kitiyanan. Shape- and texture-based fish image recognition system. In *Kasetsart J*, 2013.

[24] Andrew Rova, Greg Mori, and Lawrence M Dill. One fish, two fish, butterfish, trumpeter : Recognizing fish in underwater video. In *MVA*, pages 404–407, 2007.

[25] Jürgen Schmidhuber. Deep learning in neural networks : An overview. *Neural Networks*, 61 :85–117, 2015.

[26] Bernhard Schölkopf, Christopher JC Burges, and Alexander J Smola. *Advances in kernel methods : support vector learning*. MIT press, 1999.

[27] Yi-Haur Shiau, Sun-In Lin, Yi-Hsuan Chen, Shi-Wei Lo, and Chaur-Chin Chen. Fish observation, detection, recognition and verification in the real world. In *International Conference on Image Processing, Computer Vision and Pattern Recognition*, 2012.

[28] Concetto Spampinato, Simone Palazzo, Daniela Giordano, Isaak Kavasidis, Fang-Pang Lin, and Yun-Te Lin. Covariance based fish tracking in real-life underwater environment. In *VISAPP (2)*, pages 409–414, 2012.

[29] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[30] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2) :137–154, 2004.

[31] Qiang Zhu, M-C Yeh, Kwang-Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1491–1498. IEEE, 2006.