Examen du cours de troisième année :
Comprendre les données visuelles à grande échelle
12 février 2020

durée : 2h00

Documents autorisés : transparents imprimés du cours, notes de cours.
Téléphone portable et ordinateur non autorisés.
Il est conseillé de lire tous les exercices avant de commencer.

Nom :
Prénom :

EXERCICE 1. Questions à réponses courtes

Répondez de manière *concise*. Sauf si cela est explicitement demandé, vous n'êtes pas obligés de justifier vos réponses.

[Question 1.] Questions à choix multiples : Choisissez les propositions correctes parmi les propositions suivantes. Répondez directement sur la feuille d'énoncer.
Attention : les mauvaises réponses sont pénalisées.

(a) Le taux de faux positifs est la même chose que la précision. (i) vrai, (ii) faux.

(b) Le taux de vrais positifs est la même chose que le rappel. (i) vrai, (ii) faux.

(c) L'erreur de modélisation (*modeling error*) dans le cadre de l'apprentissage supervisé est-elle : (i) évitable, (ii) inévitable ?

(d) Un moteur de recherche est-il basé sur un algorithme de i) classification, ii) détection, iii) recherche d'images, iv) segmentation ?

(e) Le détecteur DPM contient un modèle géométrique par parties pour décrire les catégories d'objets : i) vrai, ii) faux.

(f) Dans un réseau de neurones, une couche complètement connectée (*fully connected layer*) contient des poids qui peuvent être appris au moment de l'entraînement : i) vrai, ii) faux.

(g) Dans un réseau de neurones, une couche de max-pooling (*max-pooling layer*) contient des poids qui peuvent être appris au moment de l'entraînement : i) vrai, ii) faux.

(h) Un réseau de neurones avec plusieurs couches convolutionnelles (aucun pooling, aucune non-linéarité) peut être transformé en un réseau à une seule couche : i) vrai, ii) faux.

(i) Un réseau de neurones récurrent (*recurrent neural network*) peut être appris par descente de gradient. (i) vrai, (ii) faux ?

(j) Le descripteur SIFT est invariant à la translation. (i) vrai, (ii) faux ?

[Question 2.] Lors de la création d'un système de recherche d'images, quelles sont les variations d'apparence auxquelles un bon descripteur d'images se doit d'être robuste ?

[Question 3.] Choisissez une représentation locale, c'est-à-dire choisissez un couple (détecteur, descripteur), mentionnez votre choix, et pour chacune des variations d'apparence mentionnées plus haut, dites si votre paire (détecteur, descripteur) est invariante ou non à ces variations et pourquoi.

[Question 4.] Décrivez en une phrase le descripteur visuel global R-MAC. Listez une de ses principales limitations.

[Question 5.] Décrivez en une phrase l'algorithme de l'analyse en composantes principales (ACP ou *PCA* en anglais).

[Question 6.] Expliquez succinctement le concept de couche d'inversion de gradient (*gradient reversal layer*). A quelle fin ce type de couche est-elle utilisée dans un réseau de neurones ?

[Question 7.] Qu'est ce qu'une fonction d'erreur de triplets (*triplet loss*) ? Rappelez la formule, en prenant bien soin de définir toutes les quantités. Cette fonction de coût (*loss*) dépend-elle de paramètres ? Si oui, lesquels, et quelles sont leur rôle ?

[Question 8.] Vous développez un système de vidéo-surveillance qui servira à retrouver des individus dans une base d'enregistrements vidéo à partir d'une simple photographie. A quelle mesure ferez-vous particulièrement attention : la précision ? le rappel ? Pourquoi ?

[Question 9.] Vous développez un moteur de recherche pour retrouver des photos qui seront utilisées à des fins d'illustration pour une plaquette publicitaire. A quelle mesure ferez-vous particulièrement attention : la précision ? le rappel ? Pourquoi ?

EXERCICE 2. Représentations d'images

On suppose qu'un détecteur de points d'intérêt a sélectionné trois régions dans une image, et que le descripteur choisi pour les représenter fourni les descripteurs suivants : $x_1 = (2, 1, 0)$, $x_2 = (1, 2, 0)$, $x_3 = (0, 0, 2)$.

Le vocabulaire visuel est composé de quatre mots : $A = (1, 1, 0)$, $B = (0, 0, 1)$, $C = (1, 0, 1)$, et $D = (-1, 1, 0)$.

[Question 10.] A quel mot visuel est associé chaque descripteur ?

[Question 11.] Donner la représentation par sac-de-mots (ou *bag-of-visual-words representation*) pour cette image.
Note : le détail des calculs n'étant pas demandé, un croquis pourrait éventuellement vous faire gagner du temps et éviter des calculs.

[Question 12.] Calculer la représentation VLAD pour cette image, toujours en utilisant le même vocabulaire visuel (ou *codebook*).

[Question 13.] Quelle information supplémentaire manque-t-il afin d'être en mesure de calculer la représentation par vecteur de Fisher (ou *Fisher Vector*) ?

EXERCICE 3. Système de recherche d'images

On suppose disposer d'une base de 5 images représentées chacune par un descripteur global. Les images 1 à 5 sont donc respectivement représentées par les représentations $b_1 = (1,1,0,0,1,0,1,0)$, $b_2 = (1,0,1,0,0,0,0,2)$, $b_3 = (2,1,0,0,0,0,0,0)$, $b_4 = (1,3,0,0,0,0,1,0)$, $b_5 = (0,0,1,0,0,0,1,1)$.

[Question 14.] On suppose recevoir une requête image dont la représentation est la suivante : $b_q = (2,1,0,0,0,0,0,1)$. Parmi les 5 images de la base, quelle est l'image la plus similaire à la requête ?

[Question 15.] Toujours avec la même requête, classer les images 1 à 5 de la plus similaire à la requête à la moins similaire.

[Question 16.] Les images 1, 3 et 4 appartiennent à la catégorie A, tandis que les images 2 and 5 appartiennent à la catégorie B. La requête, quant à elle, appartient à la catégorie A. En utilisant ces informations, donnez la vérité terrain de pertinence pour chacune des images de la base pour la requête donnée.

[Question 17.] Tracez la courbe précision-rappel correspondant aux résultats retournés pour la requête en utilisant la vérité terrain précédemment créée.

EXERCICE 4. Base d'entraînement et modèle pour la classification

On souhaite créer une base d'images pour entraîner un classifieur d'oiseaux capable de reconnaître à quelle espèce appartient un oiseau photographié entre plusieurs centaines d'espèces.

[Question 18.] A quoi faut-il faire attention au moment de la sélection des images qui formeront cette base d'apprentissage ? Commentez également sur le choix de la difficulté des images, de leur diversité, ou des sources utilisées pour la collection.

[Question 19.] A quoi faut-il faire attention au moment de l'annotation des images de cette base d'apprentissage ? Commentez également sur le type d'annotation que vous collecteriez.

À partir de ces données on décide d'implémenter un système de reconnaissance d'espèce d'oiseaux dans les images (100 espèces par exemple) avec un framework *end-to-end*, en utilisant un réseau de neurones convolutif, qui contient des couches convolutionelles, des couches de *pooling* et des couches de *ReLU*.

[Question 20.] Combien de couches utiliseriez-vous pour réaliser cet objectif de reconnaissance ? Détaillez votre réponse, avec l'entrée, la/les sortie(s), et les couches intermédiaires.

[Question 21.] Quels sont les paramètres de ce modèle ? Décrivez une méthode pour les apprendre.

[Question 22.] Pourrait-on utiliser un autre modèle de classification ? Détaillez votre réponse.