

## Deep Learning – Partie I – EISTI option IA

Durée : 1h30

Pendant l'examen les documents de cours, vos notes personnelles et vos ordinateurs portables ne sont pas autorisés.

Veillez répondre aux questions ci-dessous du mieux que vous pouvez.  
Les questions peuvent comporter plusieurs réponses correctes.

Nom \_\_\_\_\_

Prénom \_\_\_\_\_

## QUESTIONS

**Question 1** – Quelles sont les avantages des filtres convolutifs pour la détection d'objets ?

- Invariance aux variations dues à la rotation
- Invariance aux variations dues à la translation
- Invariance aux conditions d'illumination

**Question 2** – Dans la synthèse additive correspondant aux images numériques, quelle couleur obtient t'on en additionnant du rouge, du bleu et du vert ?

Du blanc

**Question 3** – Cochez le résultat de la convolution 1D du signal par le filtre ci-dessous :

Signal : [ 1, 1, 2, 1, 1, 0, 1, 1 ]

Filtre : [1, 0, 1]

- [2, 2, 4, 2, 2, 0, 2, 2]
- [3, 2, 3, 1, 2, 1]
- [-1, 0, 1, 1, 0, -1]
- [0, 0, 2, 0, 0]

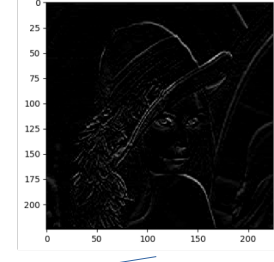
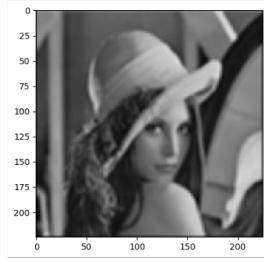
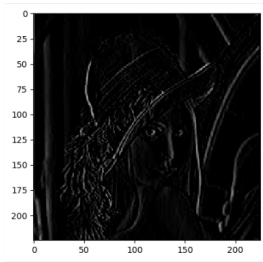
**Question 4** – Quel est l'intérêt d'un filtre gaussien par rapport à un simple filtre moyenneur ?

Filtre gaussien : [0.18, 0.21, 0.26, 0.21, 0.18]

Filtre moyenneur : [0.2, 0.2, 0.2, 0.2, 0.2]

- Accorder plus d'importance aux pixels correspondant au centre du filtre par rapport à ceux de la périphérie.
- Des routines optimisées en C permettent de gérer les calculs gaussiens plus efficacement.
- Être plus robuste aux variations dues à l'illumination

**Question 5** – Reliez par un trait chaque matrice de filtre avec l’image qu’elle a produit.



$$\begin{pmatrix} 0.03 & 0.03 & 0.4 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.04 & 0.037 \\ 0.03 & 0.04 & 0.05 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.04 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$(1 \ 0 \ -1)$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

**Question 6** – Quelle est la raison de l’utilisation du padding ?

Dans le cas des filtres convolutifs classiques, il s’agit de gérer les effets de bords sur l’image.

D’autres utilisations en sont faites, par exemple, dans les convolutions atrous.

**Question 7** – Pour quelle(s) raison(s) la taille du résultat d’une convolution peut être différente par rapport à la taille de l’image originale ?

- Utilisation du padding
- Utilisation d’un pas de convolution
- Taille du filtre
- Valeurs des coefficients du filtre

**Question 8** – Certaines techniques d’augmentation de la résolution des images utilisent une convolution ?

Vrai

**Question 9** – En machine learning, une déconvolution, c.a.d. l’inverse de la convolution par un filtre, peut être effectuée par :

- x Une convolution en utilisant le même filtre en sens inverse et en insérant des zéros dans le signal initial.
- o Une convolution en utilisant un filtre dont les coefficients sont l'inverse des coefficients du filtre original.
- o Il est impossible de déconvoluer exactement un signal

**Question 10** – L'intérêt d'inclure la géométrie dans le Deep Learning peut être :

- x L'utilisation d'un signal non supervisé pour guider l'apprentissage.
- x Inclure dans l'apprentissage des aprioris sur la structure du monde.
- o Inclure des informations sémantiques (e.g. sur le sens que les humains donnent aux objets) dans l'apprentissage.

**Question 11** – Relier chaque image à sa transformation affine correspondante par un trait.



$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{pmatrix}$$

**Question 12** – En géométrie, la principale source d'erreur vient des mesures bruitées. Elles correspondent à :

- o Des imprécisions numériques lors du calcul de la pseudo-inverse.
- x Des imprécisions dans le calcul des correspondances de points entre les images.
- o Le modèle de caméra perspective n'est pas assez conforme à la réalité.

**Question 13** – L'un des inconvénients majeurs de la fonction de coût à base des moindres carrés est :

- x Elle est sensible aux points aberrants générés par des mesures bruitées.
- o Elle devient complexe à gérer quand le rapport nombre d'équations / nombre d'inconnues augmente.
- o La méthode de la pseudo inverse n'est pas équivalente à sa réduction.

**Question 14** – Donnez le sens des paramètres de la caméra perspective P :

$$P = \begin{pmatrix} f \times \lambda_x & 0 & \sigma_x \\ 0 & f \times \lambda_y & \sigma_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix}$$

f : focale : distance entre le plan de l'image

lambda : Conversion unité dans le repère du monde => unités des pixels.

sigma : Centre de l'image en pixels.

R : orientatio de la caméra

T : Translation de la caméra dans le repère du monde.

**Question 16** – Les techniques de « structure à partir du mouvement » comme la méthode de Tomasi et Kanade permettent de :

x Estimer les matrices de caméras et la position 3D des points à partir de correspondances entre pixels.

o Estimer une transformation à partir de deux images.

o Estimer les matrices de caméras étant données des coordonnées de points en 3D.

o Estimer la position de points en 3D étant données les matrices de caméras et la position de ces points sur les images.

### Résolution de projet :

On vous a chargé d'installer un système de réalité augmentée dans un musée avec des lunettes équipées de caméras qui sont fournies à chaque visiteur. Pour connaître les éléments virtuels à afficher, votre système doit estimer la localisation du porteur des lunettes. Cette estimation peut être formulée comme un problème de « Simultaneous Localization and Mapping » (SLAM), où, étant données les images d'une caméra se déplaçant dans l'espace, le système estime au fur et à mesure de sa progression la position de la caméra tout en construisant une carte de l'espace environnant en 3D. C'est en fait le même problème que le « Structure à partir du mouvement », à la différence que les images ne sont fournies qu'une par une.

*On suppose que l'on connaît les paramètres internes des caméras ( focale, etc). Quelles sont les paramètres qui restent à estimer ?*

L'orientation et la translation.

Dans le contexte du SLAM, effectuer la reconstruction avec toutes les images fournis par les caméras est couteux en temps de calcul, surtout si l'on veut une mise à jour en temps réel. Cependant, toutes les images ne sont pas utiles à la reconstruction.

*- Quelles sont les mouvements de caméras dont vous pouvez ignorer les images dans votre système ?*

*x Rotation de la caméra sur son axe*

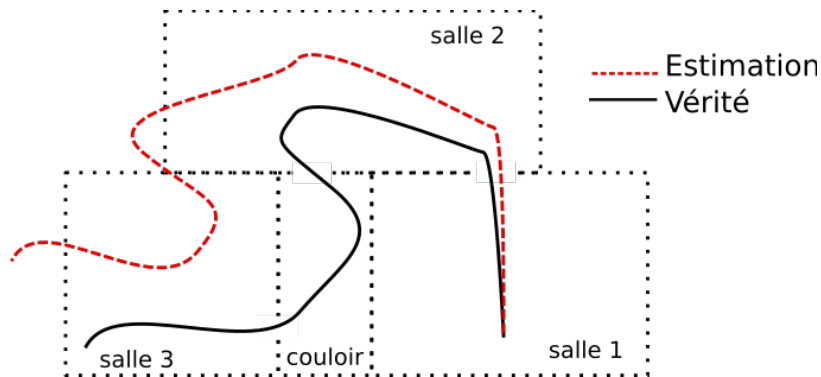
*x Caméra statique sans mouvement*

*o Translation*

*- Intuitivement, quelles sont les conditions idéales du mouvement de la caméra favorisant la reconstruction ?*

Un mouvement de translation permettant d'observer de nombreux points sous des angles les plus grands possibles.

Lors de l'estimation de la reconstruction 3D et de la trajectoire de la caméra en ligne, la position au temps  $t+1$  est estimée à partir de la position du temps  $t-1$ . Cette configuration a l'inconvénient que les erreurs s'accumulent au cours du temps. On parle alors de «drift» comme illustré sur la figure ci-contre.



*Comment améliorer l'estimation quand la caméra repasse sur une scène qui a déjà été observée ?*

Dés qu'un point est visible sous plusieurs angles, il est possible d'inclure cette information en appliquant une méthode de type « Structure from motion » .

Une autre solution au problème de drift est de placer des repères visuels dans le musée qui seront reconnues par la caméra. *Détaillez en quelques lignes comment vous implémenteriez cette solution et de quelles informations vous auriez besoin.*

Des repères visuels, comme des QR codes pourraient être placés sur les murs. La position en 3D de ces repères dans le repère du monde sera mesurée au préalable. Ces points peuvent servir alors de référence pour améliorer la précision.