

# ROB 313 - IA 323

## Vision pour la Robotique

## Contrôle des Connaissances

28 Janvier 2022

### 1 Caméras RGB-d par lumière structurée

Quelles propriétés doit respecter la mire projetée dans une scène par le projecteur d'une caméra RGB-d active à lumière structurée? Donner deux exemples de mires possibles. Quelles sont les difficultés et limites de ces caméras?

### 2 Contrainte épipolaire

Qu'est-ce que la contrainte épipolaire qui lie 2 images de la même scène perçue depuis deux points de vue différents? Comment s'exprime-t-elle mathématiquement? Lorsque les deux points de vue sont fournis par la même caméra qui s'est déplacée entre deux instants  $t_1$  et  $t_2$ , comment peut-on utiliser cette contrainte pour détecter les objets qui se sont déplacés par rapport à la scène entre  $t_1$  et  $t_2$ ?

### 3 Détection 3d monoculaire de personnes

Soit une caméra (fixe) observant le grand hall de l'ENSTA. Supposons qu'on applique une détection de personnes à chaque image qui fournit les coordonnées 2d de la boîte englobant chaque personne qui se trouve dans le grand hall. Comment peut-on estimer la position 3d de la personne, c'est-à-dire sa distance à la caméra, et les dimensions réelles (i.e. en cm) de la boîte englobante?

### 4 Bag-of-features

Un robot ménager muni d'une caméra couleur a été entraîné pour reconnaître la pièce dans laquelle il se trouve à partir de son apparence visuelle, en utilisant une approche de type bag-of-features. Décrivez une procédure possible d'apprentissage pour le robot, ainsi qu'une méthode de prédiction.

### 5 Homographies

Qu'est-ce qu'une homographie? Dans quel(s) conditions un couple d'images est-il relié par une homographie? Peut-on estimer, selon le cas, la profondeur par appariement de points?

## 6 Prédiction monoculaire auto-supervisée de cartes de profondeur

Comment se calcule la fonction de perte/coût photométrique permettant d'entraîner des réseaux profonds à prédire des cartes de profondeur de façon auto-supervisée ? Quelles sont les contraintes et difficultés potentielles ?

## 7 Réseau de neurones : entraînement

Vous entraînez un réseau de neurones (CNN) avec 100 couches, sur une tâche de classification binaire, en utilisant une activation sigmoïde dans la couche finale, et tangente hyperbolique (tanh) pour toutes les autres couches. Vous remarquez que les poids du réseau de neurones cessent de se mettre à jour après la première epoch d'entraînement, même si votre réseau n'a pas encore convergé. Une analyse plus approfondie révèle que les gradients de ces couches sont complètement ou presque complètement nuls très tôt dans la formation. Laquelle des corrections suivantes pourrait vous aider ? (Notez également que votre fonction de perte/coût ne diverge pas).

1. Augmenter la taille de votre ensemble d'entraînement
2. Changer les fonctions d'activation tanh pour des ReLU
3. Ajouter de la batch normalisation avant chaque activation
4. Augmenter le learning rate

## 8 Réseau de neurones : fonctions d'activation

Parmi les fonctions suivantes, lesquelles considérez-vous comme des fonctions d'activation valides (non-linéarités élément par élément) pour entraîner un réseau de neurones dans la pratique ?

1.  $f(x) = -\min(2, x)$
2.  $f(x) = 0.9x + 1$
3.  $f(x) = \begin{cases} \min(x, .1x) & \text{if } x \geq 0 \\ \min(x, .1x) & \text{if } x < 0 \end{cases}$
4.  $f(x) = \begin{cases} \max(x, .1x) & \text{if } x \geq 0 \\ \min(x, .1x) & \text{if } x < 0 \end{cases}$

## 9 Réseau de neurones : Tracking

Supposons que vous souhaitiez entraîner un réseau de neurones (CNN) pour suivre les particules en biologie, nous considérons que vous disposez d'un jeu de données composé d'images de cellule et de leur emplacement correspondant. Quel type d'architecture d'apprentissage en profondeur pouvez-vous utiliser pour suivre la particule :

1. faster RCNN
2. DeepLab
3. Unet
4. Yolo

## 10 Réseau de neurones : cartes de profondeur

Supposons que vous souhaitiez un réseau de neurones pour prédire des cartes de profondeur (3D), l'entrée étant composée d'une seule image. Ce type de CNN prédit une profondeur pour chaque pixel. Quel type d'architecture de CNN pouvez-vous utiliser après une modification mineure pour prédire la profondeur monoculaire :

1. faster RCNN
2. DeepLab
3. Unet
4. Yolo

## 11 Réseau de neurones : cartes de profondeur

Supposons que vous souhaitiez entraîner un réseau de neurones à prédire des cartes de profondeur (3D). Ce type de CNN prédit une profondeur pour chaque pixel. Notons  $y \in \mathbb{R}^K$  la vérité terrain et  $\hat{y} \in \mathbb{R}^K$  la prédiction du CNN qui sont toutes les deux des cartes de profondeur de taille  $K$ . Quel type de fonction de perte/coût pouvez-vous utiliser pour entraîner votre réseau :

1. Cross entropie :  $H(y, \hat{y}) = - \sum_{i=1}^K y_i \log(\hat{y}_i)$
2. Mean Squared Error :  $MSE(y, \hat{y}) = \frac{1}{K} \sum_{i=1}^K (y_i - \hat{y}_i)^2$
3. Mean Absolute Error :  $MAE(y, \hat{y}) = \frac{1}{K} \sum_{i=1}^K |y_i - \hat{y}_i|$
4. Divergence de Kullback Leibler :  $D_{KL}(y, \hat{y}) = \sum_{i=1}^K y_i \log(\frac{y_i}{\hat{y}_i})$