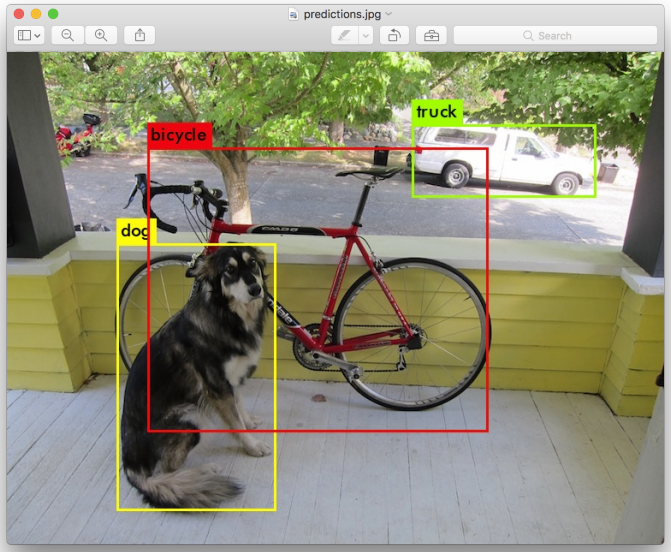
Quelques idées sur la reconnaissance d'objets

Marie-Odile Berger, INRIA Nancy Grand Est

February 22, 2021

Reconnaissance et détection d'objets

- reconnaissance: présence d'une instance de l'objet (« une voiture, un chat ») dans l'image
- détection: situation précise de l'instance dans l'image (cadre englobant ou segmentation de la zone)
- carte sémantique: chaque pixel est étiqueté par une classe (route, ligne, panneau, arbre...)



De la reconnaissance à la détection

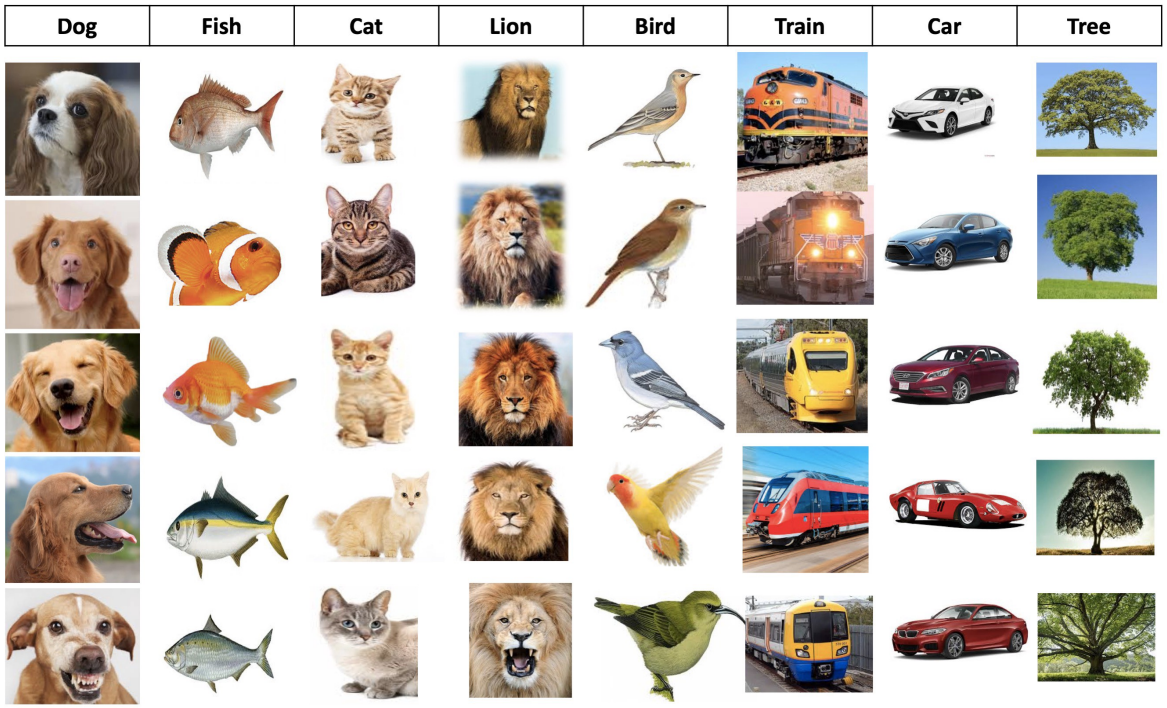
La méthode des fenetres glissantes (sliding widows):

- Faire glisser une fenetre sur toute l'image
- Reconnaitre si un objet est présent dans chaque fenetre
- Utiliser la technique avec des tailles de fenetres différentes pour identifier l'objet avec différentes tailles

[Start Video]

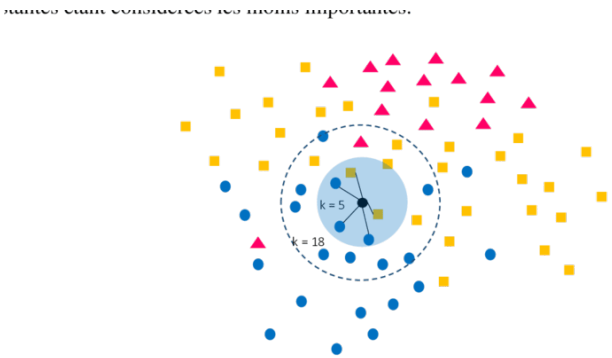
La reconnaissance par apprentissage

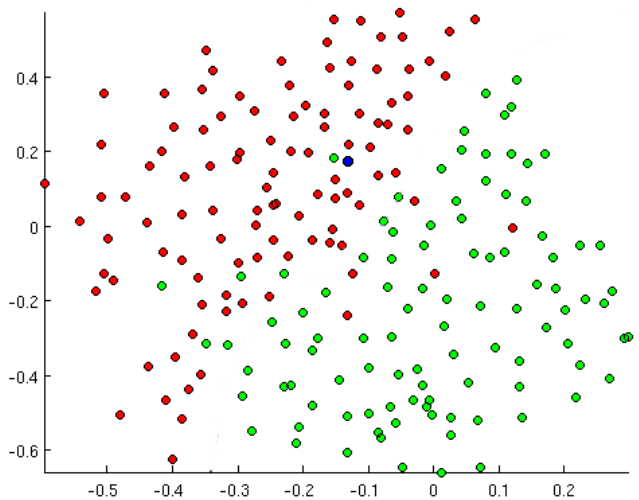
- Objectif: attribuer une classe à une image (voituren chat,..)
- Données: des exemples multiples (millions parfois..) de chaque classe (données supervisées: N couples ¡entrée-résultat¿
- méthodes étudiées: kNN (k-nearest neighbors), SVM et réseaux de neurones



Les méthodes k-NN

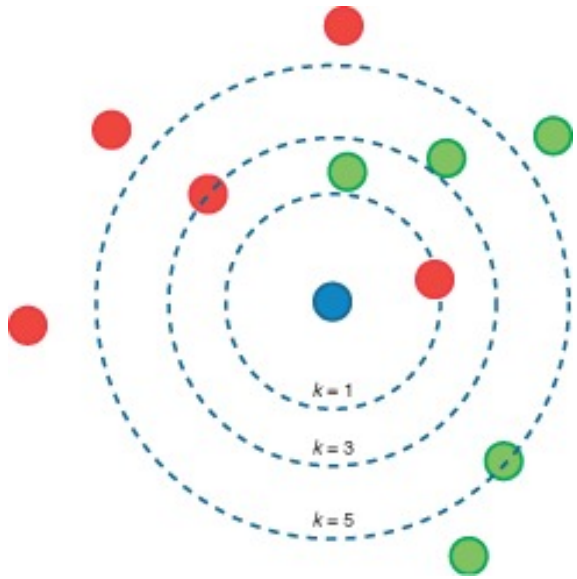
- la méthode des k plus proches voisins est une méthode d'apprentissage supervisé.
- Etant donné un élémént à classifier x, on détermine ses k plus proches voisins (k-NN). La classe de x est la classe majoritaire d'appartenance de ces k-voisins
- On se sert des données d'apprentissage, mais il n'y a pas d'autres opération que la recherche des voisins

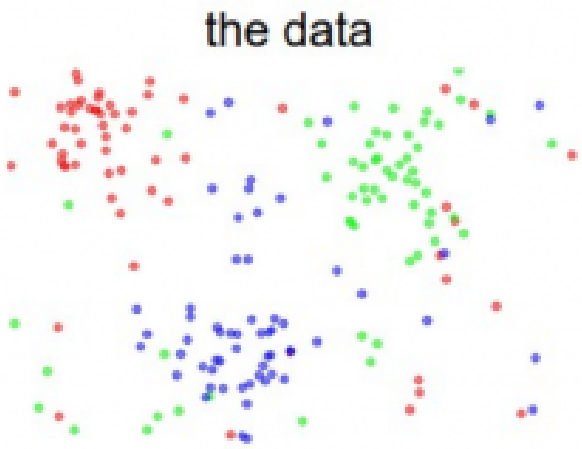


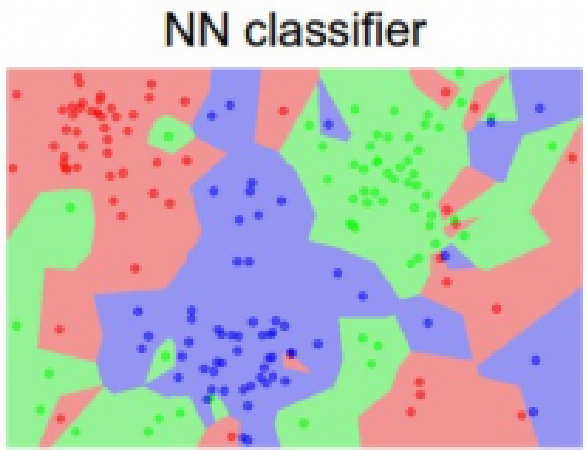


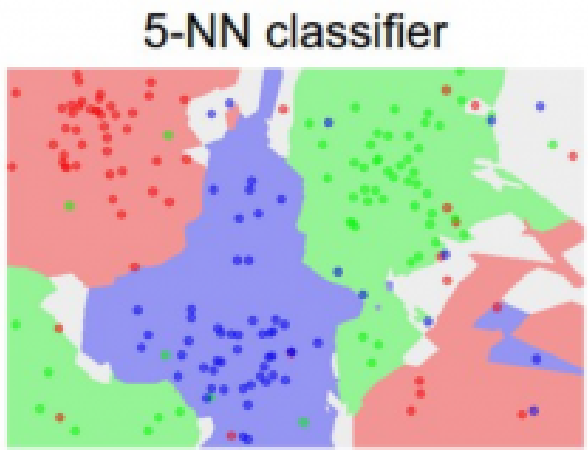
Classifieur k-NN

Difficulté: choix de k?



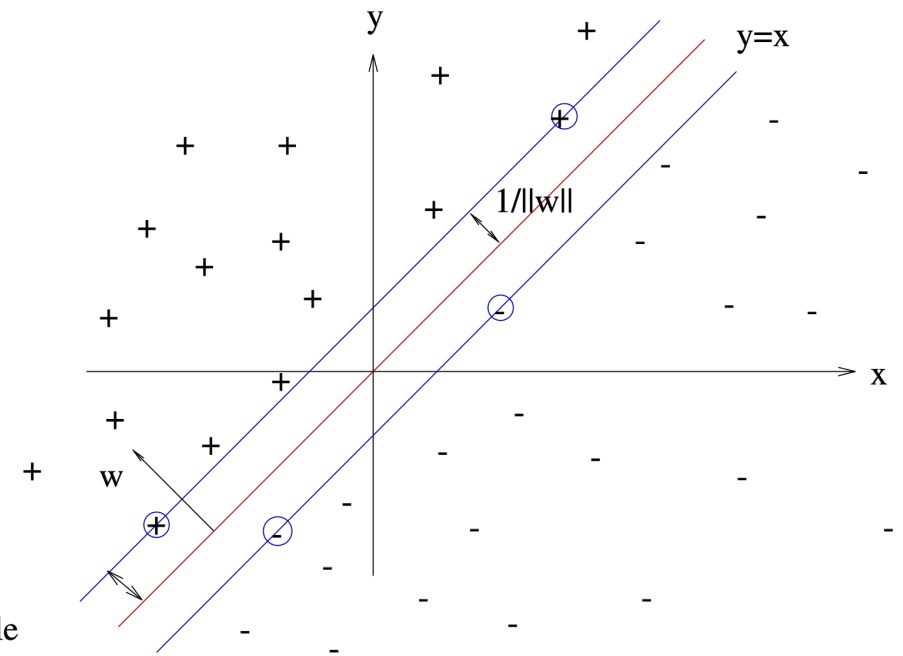






SVM: machines à vecteur support

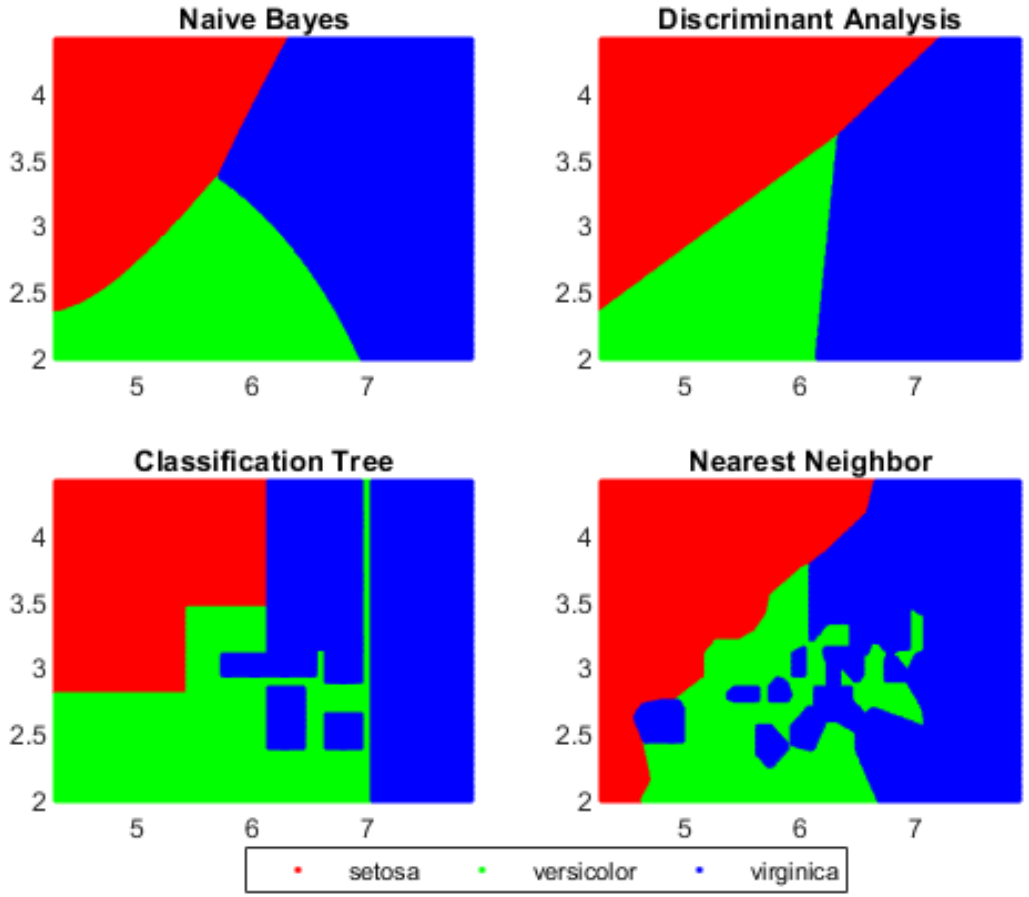
- Principe: Déterminer des hyperplan séparateurs $h(x) = \sum w_i w_x - w_0 = 0$
- x est de classe $l = 1$ si $h(x) \geq 0$, et de classe $l = -1$ si $h(x) \neq 0$. On a $l_k * (w.x - w_0) \geq 0$.
- la fonction h(x) (c'est à dire les w_i est apprise grâce à l' ensemble d'apprentissage
- Pour classifier un nouvel élément X, on regarde le signe de $h(X)$



Marge maximale

Comparaison des séparateurs issus de différentes méthodes

Voir l'exemple matlab: Visualize Decision Surfaces of Different Classifiers



Réseaux convolutionnels versus approches traditionnelles

l'approche par réseaux convolutionnels (CNN): extraction des caractéristiques et entrainement du classifieur **ne sont pas dissociées**:

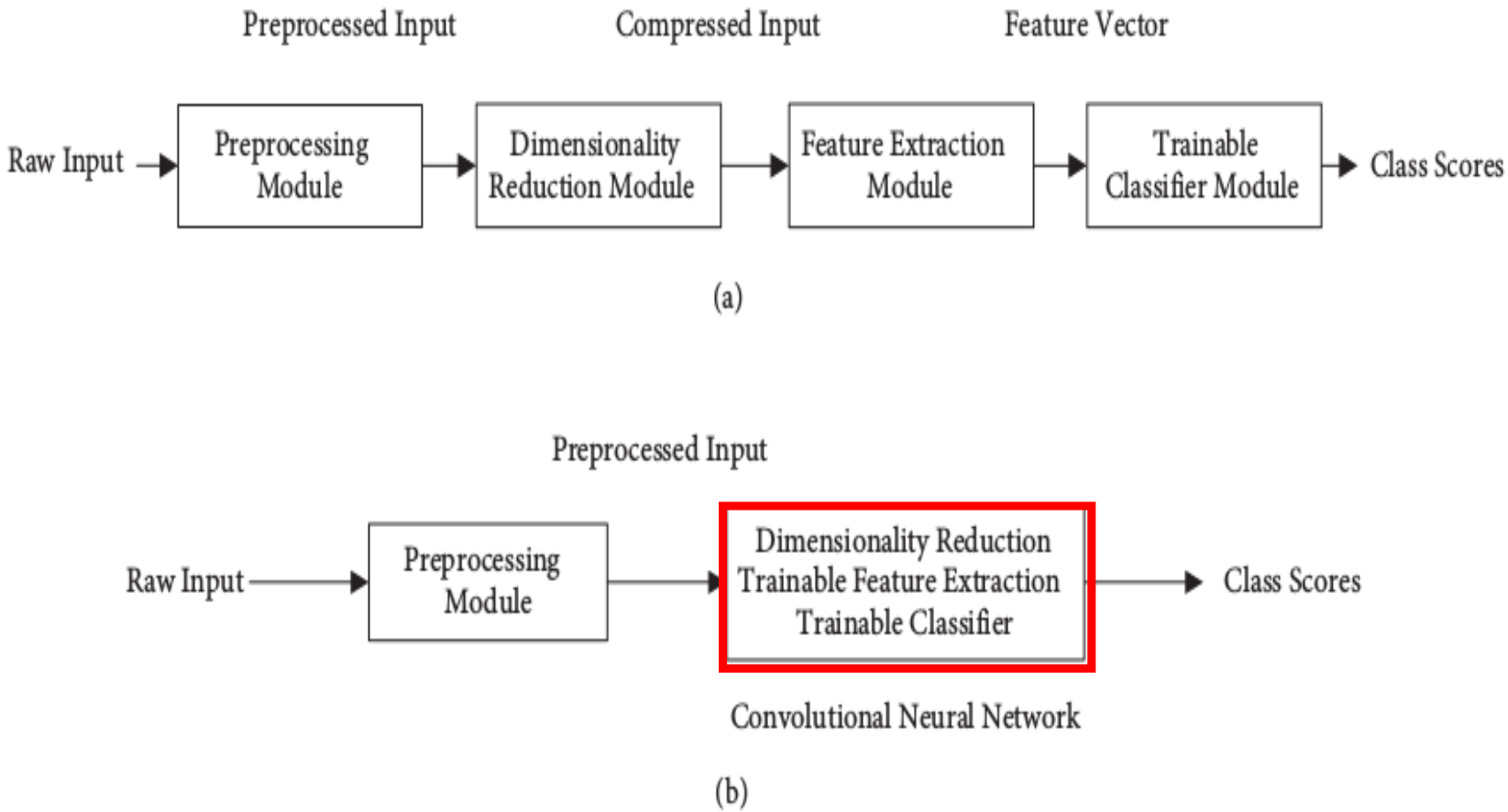


Figure 1. Pattern recognition approaches: (a) conventional, (b) CNN-based.

Réseaux de neurones

Exemple d'un des premiers réseaux convolutionnels pour la classification des chiffres: [Lecun98]

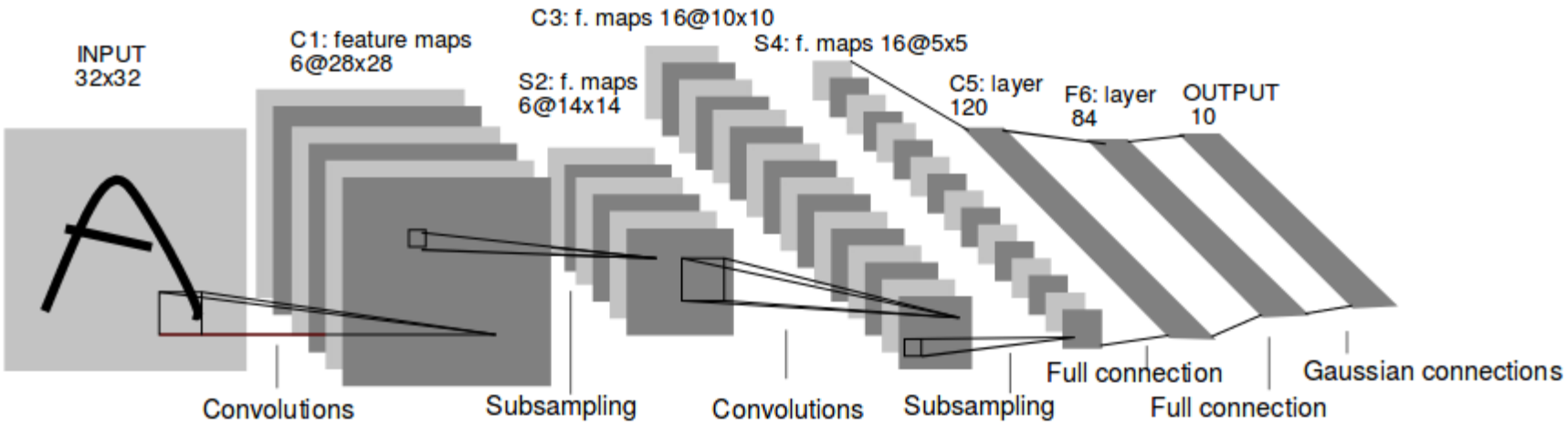


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

- Apprentissage supervisé: une liste de données Z^p dont la classe est connue D^p (vérité terrain)
- Soit W les paramètres ajustables du système (convolution,...)
- le réseau calcule une fonction $F(Z, W)$ qui fournit l'étiquette de Z

Les couches d'un réseau convolutionnel

Une succession de couches: on pourra consulter cette page de Stanford qui est un pense bête sur la structure des réseaux de neurones convolutionnels

- couches de convolution pour extraire des caractéristiques locales des images, à différentes échelles
- couche de pooling: réduire la taille tout en préservant les informations les plus importantes (garder la valeur maximale d'une fenêtre 4x4)
- couches Relu
- dernière couche: couche entièrement connectée (FC):Chaque valeur du tableau en entrée "vote" en faveur d'une classe. Les votes n'ont pas tous la même importance : la couche leur accorde des poids qui dépendent de l'élément du tableau et de la classe.

Comment choisir les paramètres de toutes ces couches?: c'est l'objectif de l'apprentissage

Réseaux de neurones et algorithme du gradient stochastique

Calcul de W?

- Les paramètres W du système sont "appris" en minimisant $E_{train}(W) = \sum dist(D^p, F(Z^p, W)$
- Minimisation par une descente de gradient... mais nombre de données d'entrée très important $\rightarrow$ le calcul du gradient est très couteux
- Recours aux méthodes de **descente de gradient stochastique**: un gradient "bruité" est évalué sur la base d'un seul exemple tiré aléatoirement (ou d'un petit lot) qui permet la re-estimation des paramètres.

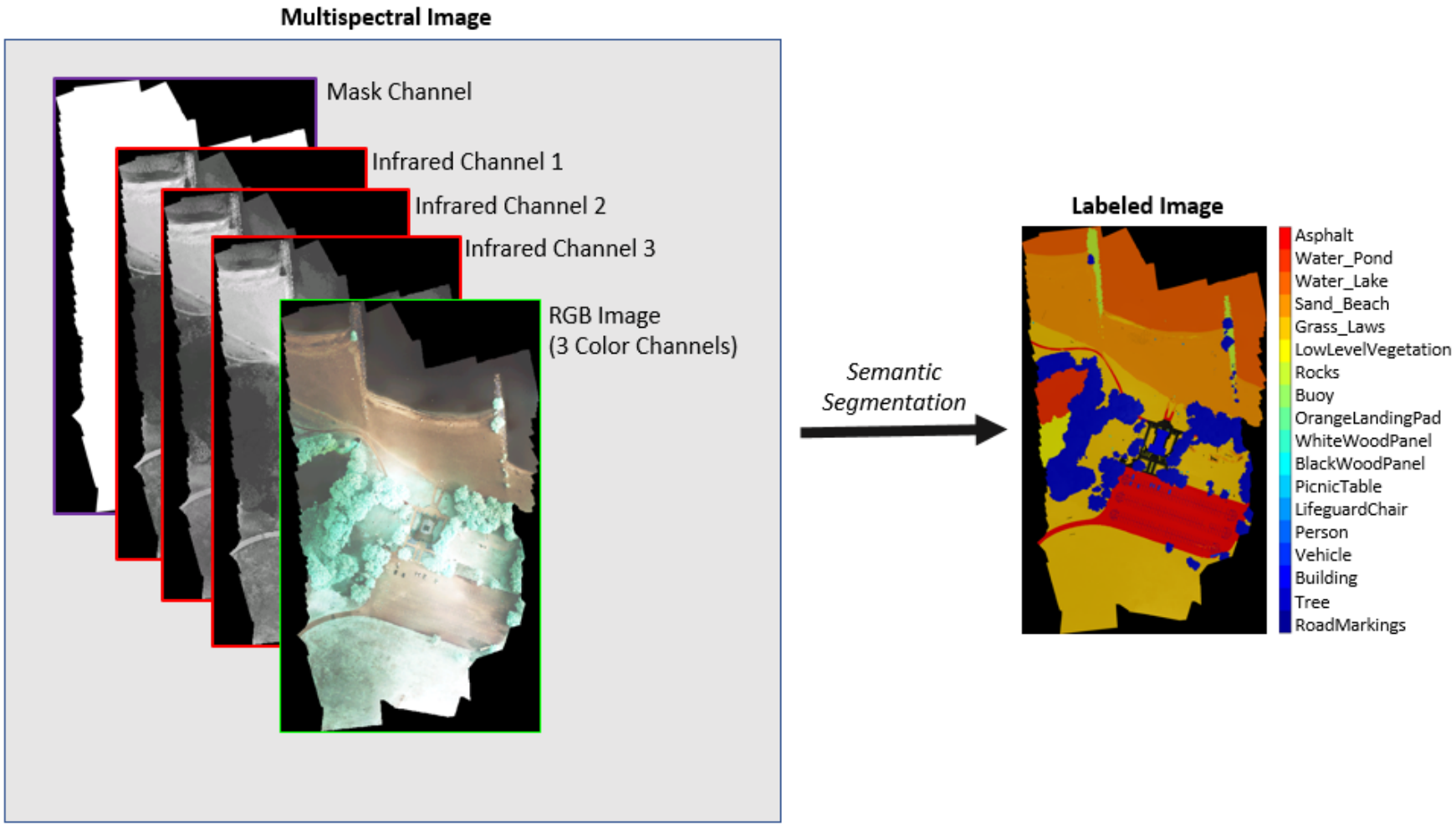
Dépendance de la base d'apprentissage!

- Toutes ces méthodes dépendent fortement de la base d'exemples utilisée pour l'apprentissage: une classe moins représentée ne sera pas (ou mal) reconnue.
- consulter l'article 2018 de Sciences et Avenir: Intelligence artificielle : la reconnaissance faciale est-elle misogyne et raciste ?
- des classifieurs *génériques* proviennent de réseaux appris sur des grandes base de données Imagenet Large Scale Visual recognition challenge. Les réseaux Alexnet, googleLeNet... sont des réseaux couramment utilisés pour la classification
- concernant les images multispectrales, nous considérerons deux travaux
  - Classification sémantique des images multispectrales (exemple dans Matlab)
  - Classification sémantiques de mesh 3D pour l'analyse de scènes urbaines

Classification sémantique d'images hyperspectrales

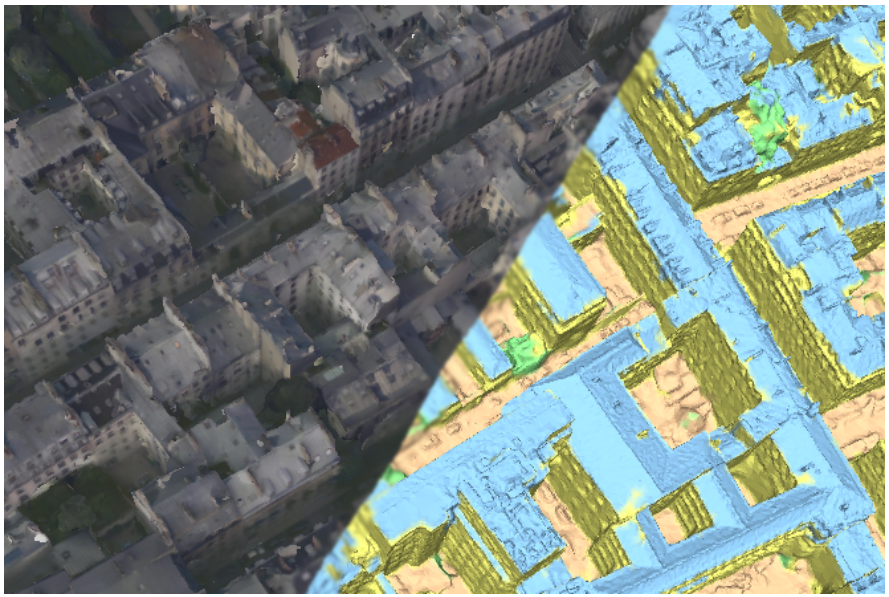
- Bases de données annotées de petites tailles par rapport aux images RGB classiques
- Cas des données décrites dans High Resolution multispectrales datasets for selantic segmentation
  - Des données *train* et *test* prises à des endroits différents à Hamlin beach state park.
  - des mosaiques d'images assemblées à partir d'images de drones et recollées par la mise en correspondance de points SIFT (voir suite du cours)
  - apprentissage sur des images alétoirement sélectionnées dans l'image train
- Etude de l'article Algorithms for Semantic Segmentation of MultispectralRemote Sensing Imagery using Deep Learning et dont l'implantation est disponible dans Matlab: voir la demo Semantic Segmentation of Multispectral Images Using Deep Learning

Classification sémantique d'images hyperspectrales



Segmentation semantique de maillages 3D

- Rouhani, Lafarge, Ailliez: Semantic segmentation of 3D textured meshes for urban scene Analysis (voir article sur ma page web)
- objectif: étiqueter chaque point d'un nuage de points 3D avec une classe en utilisant l'apprentissage.
- vecteur utilisé pour la classif: concaténation d'informations géométrique **et** phototométrique sur des super-facettes extraites des données



[Start Video]. Voir aussi la page web:ici

Précision de classification versus taille de la base d'apprentissage

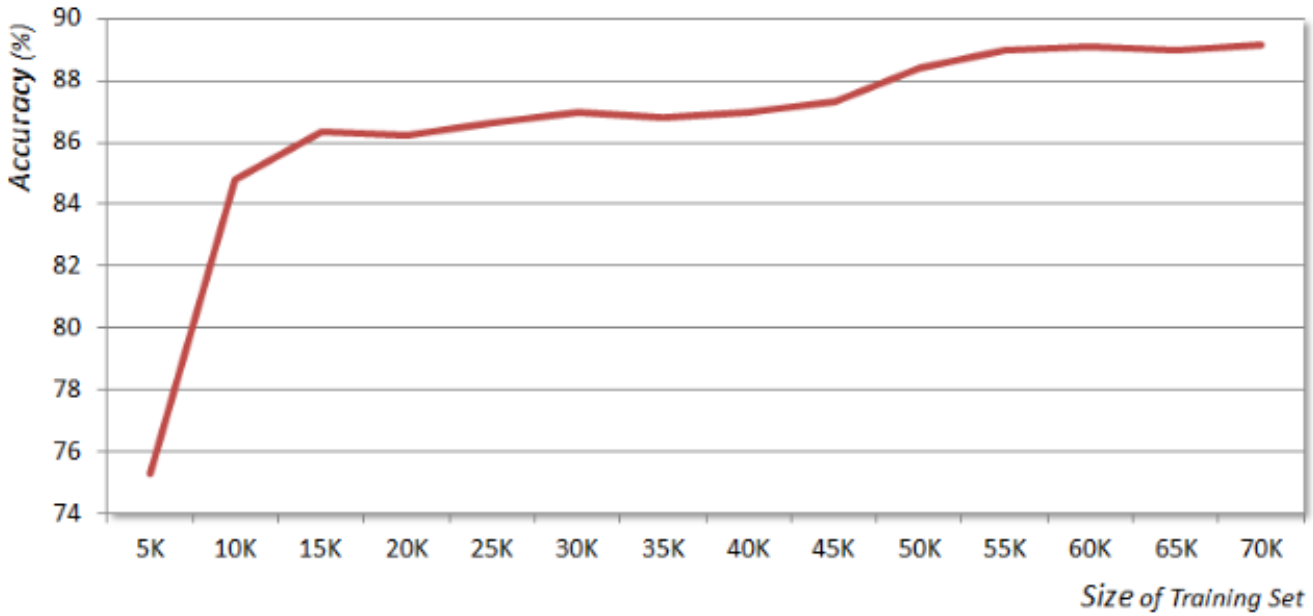


Figure 9: Classifier accuracy vs. size of training set. Using at least 10K superfacets in the training set allows the classifier to yield correct predictions for more than 85% of samples. Beyond 10K the gain obtained by increasing the number of samples is negligible.

Influence de la taille des super-facettes

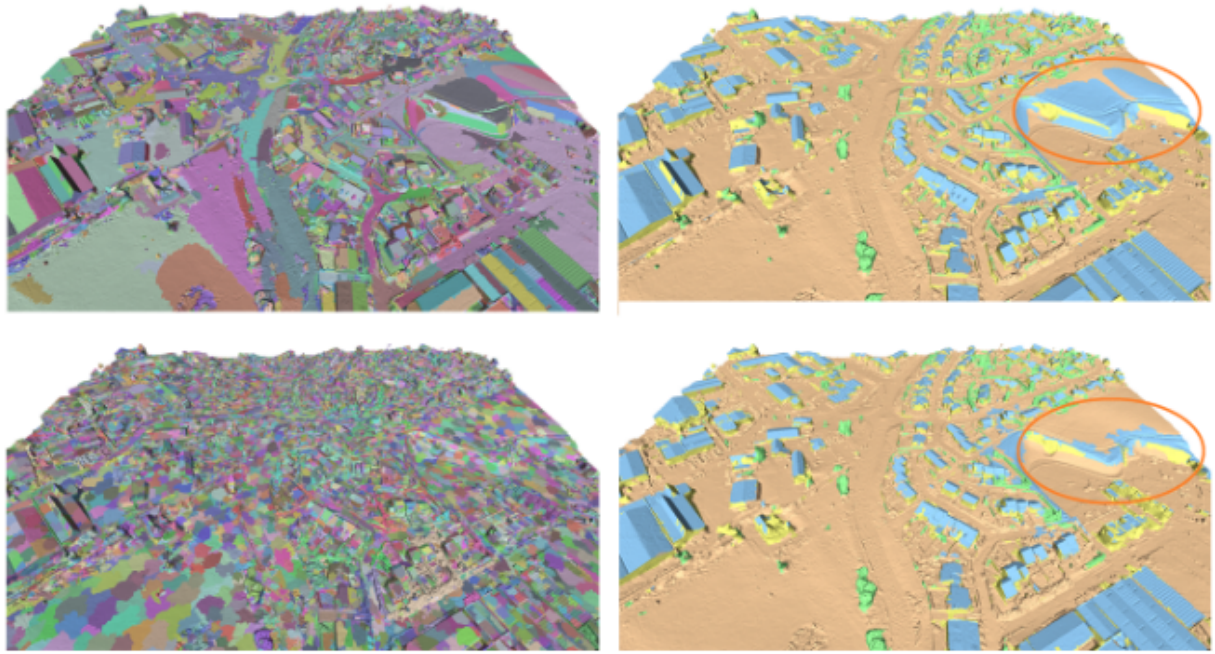


Figure 8: Size of superfacets. The use of large superfacets (top) enables fast processing. Conversely, small superfacets (bottom) often yield fewer mislabeling errors as the label propagation process operates at a more local scale. Notice the mislabeling on the hilly area.

Exemples de classification

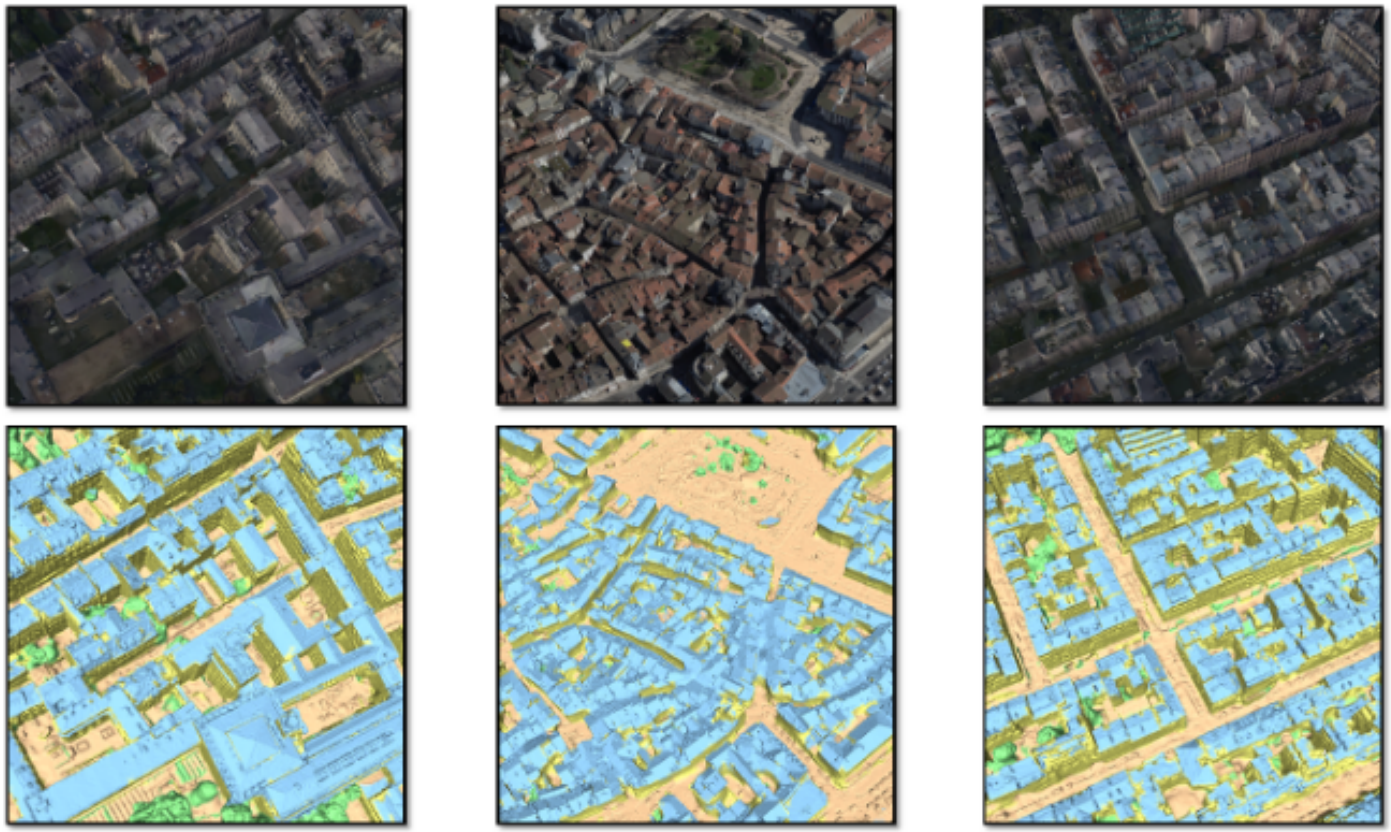


Figure 13: Classification of dense urban areas. In such urban landscapes, the main challenge is to distinguish trees from roofs and to identify ground in presence of narrow streets (see middle example). Our results are mostly correct except a few confusions for trees in inner courtyards.

Cas difficiles

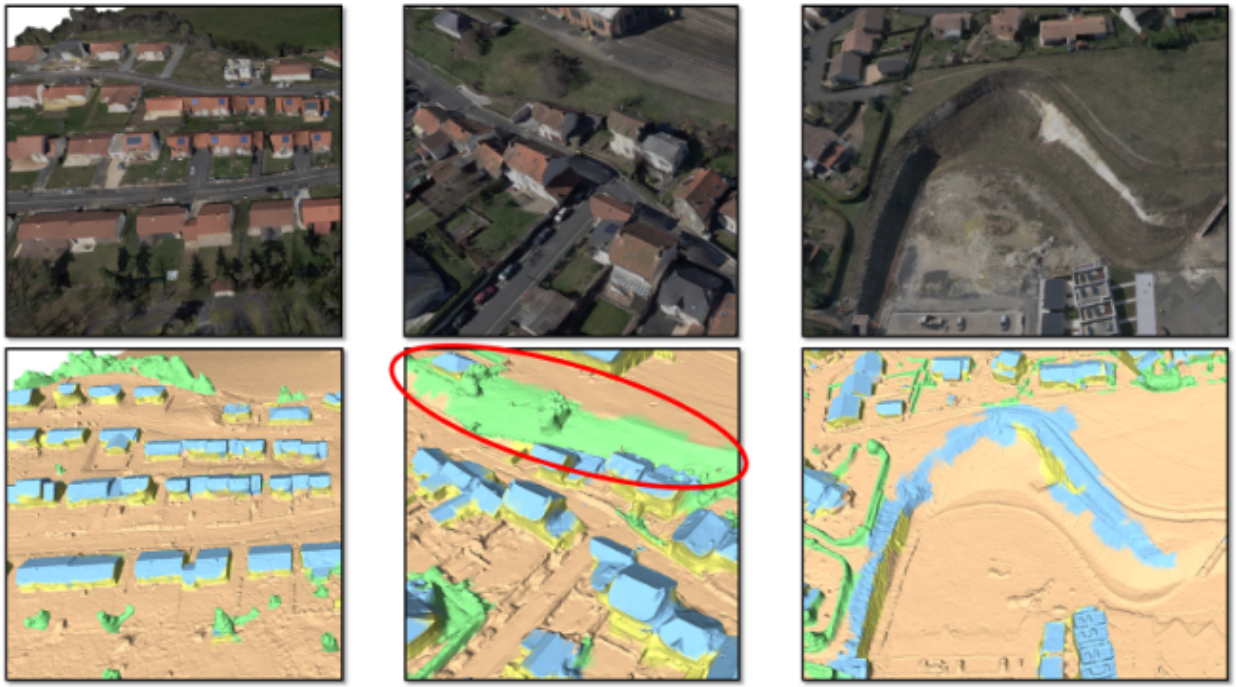


Figure 14: Classification of hilly areas. Using the elevation feature computed at different scales our classifier correctly detects sloppy ground in most cases (see left and middle). In presence of very abrupt relief changes, however, it tends to misclassify ground into facade or roof (right).