

Intelligence artificielle et données géographiques

Clément Gorin

clement.gorin@univ-paris1.fr

Université Paris 1 Panthéon-Sorbonne

Journées de formation de la CNFPT

École d'Économie de Saint-Étienne, 10 juin 2024

Introduction

Intérêt croissant de la part des chercheurs et des praticiens pour l'usage de l'apprentissage automatique

- Prévisions précises des variables économiques et des séries temporelles (Mullainathan and Spiess 2017)
- Outils non paramétriques pour l'estimation (Belloni et al. 2012) et l'inférence (Athey and Imbens 2016)
- Analyse à l'échelle de données précédemment inaccessibles (LeCun et al. 2015) ex. images, texte

De nombreuses sources d'information sont disponibles, mais encodées dans un format difficile à exploiter

- Les données sont en haute dimension et peuvent ne pas être structurées sous forme de tableau
- Les méthodes traditionnelles échouent ou donnent de mauvais résultats (Kleinberg et al. 2015)
- Le processus générateur de données est inconnu et hautement complexe (Natekin and Knoll 2013)

Ils s'agit de transformer ces structures de haute dimension en représentations simplifiées pour l'analyse

- Les méthodes d'apprentissage supervisé construisent un modèle directement à partir des données
- Formes fonctionnelles flexibles et généralisables qui ne sont pas définies à l'avance ("apprentissage")
- Approche adaptée quand la qualité des prédictions est plus importante que l'interprétabilité du modèle

1. Applications

2. Approche

3. Modèles

4. Illustrations

Applications

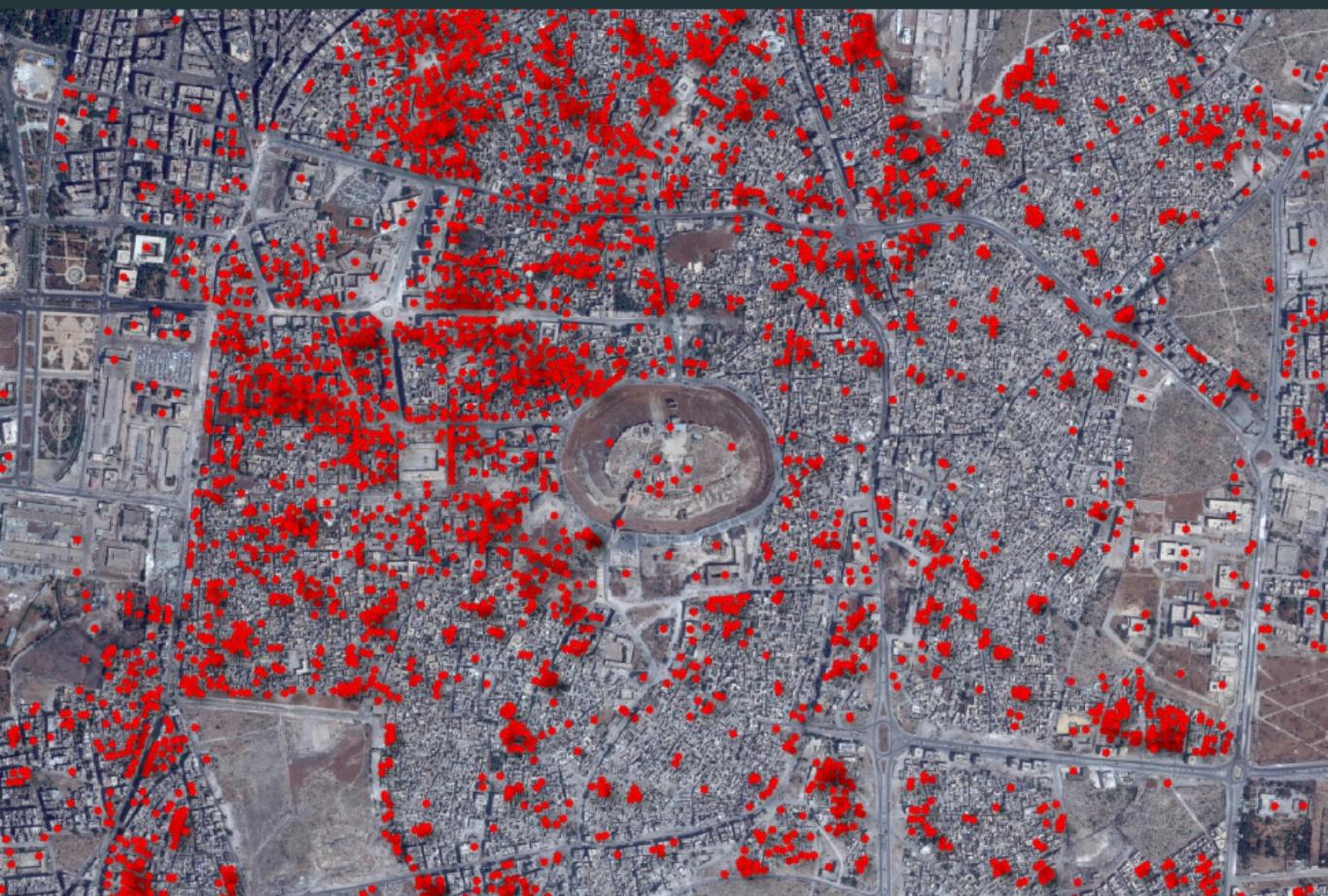
Les données spatiales sont souvent représentées sous forme de rasters, c'est à dire de matrice géolocalisée

- L'espace est discrétement à l'aide d'une grille uniforme
- Les valeurs représentent une moyenne ou une majorité du phénomène mesuré
- Informations spatiales: origine, emprise, résolution, système de référence



Applications

Données



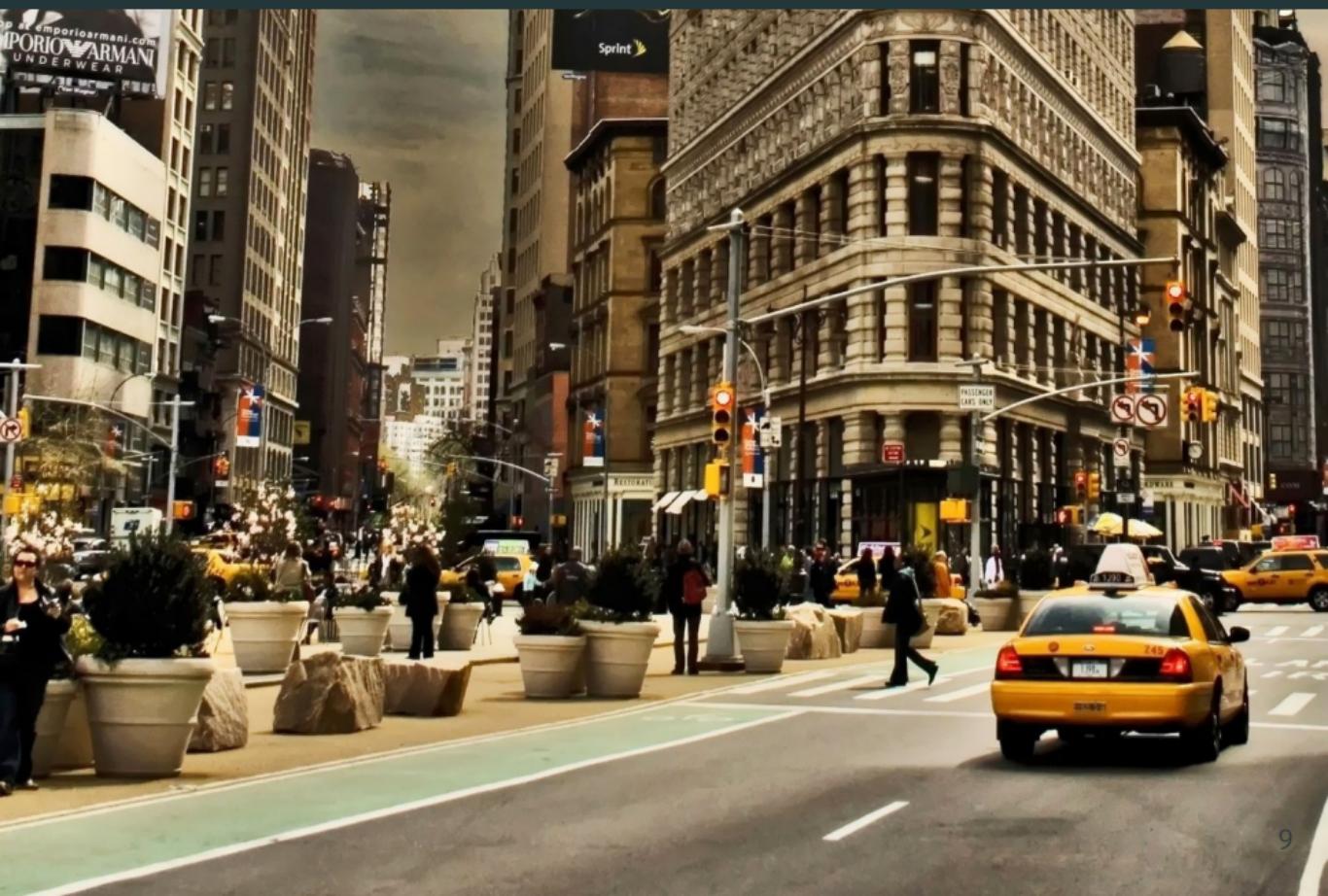
Applications

Données

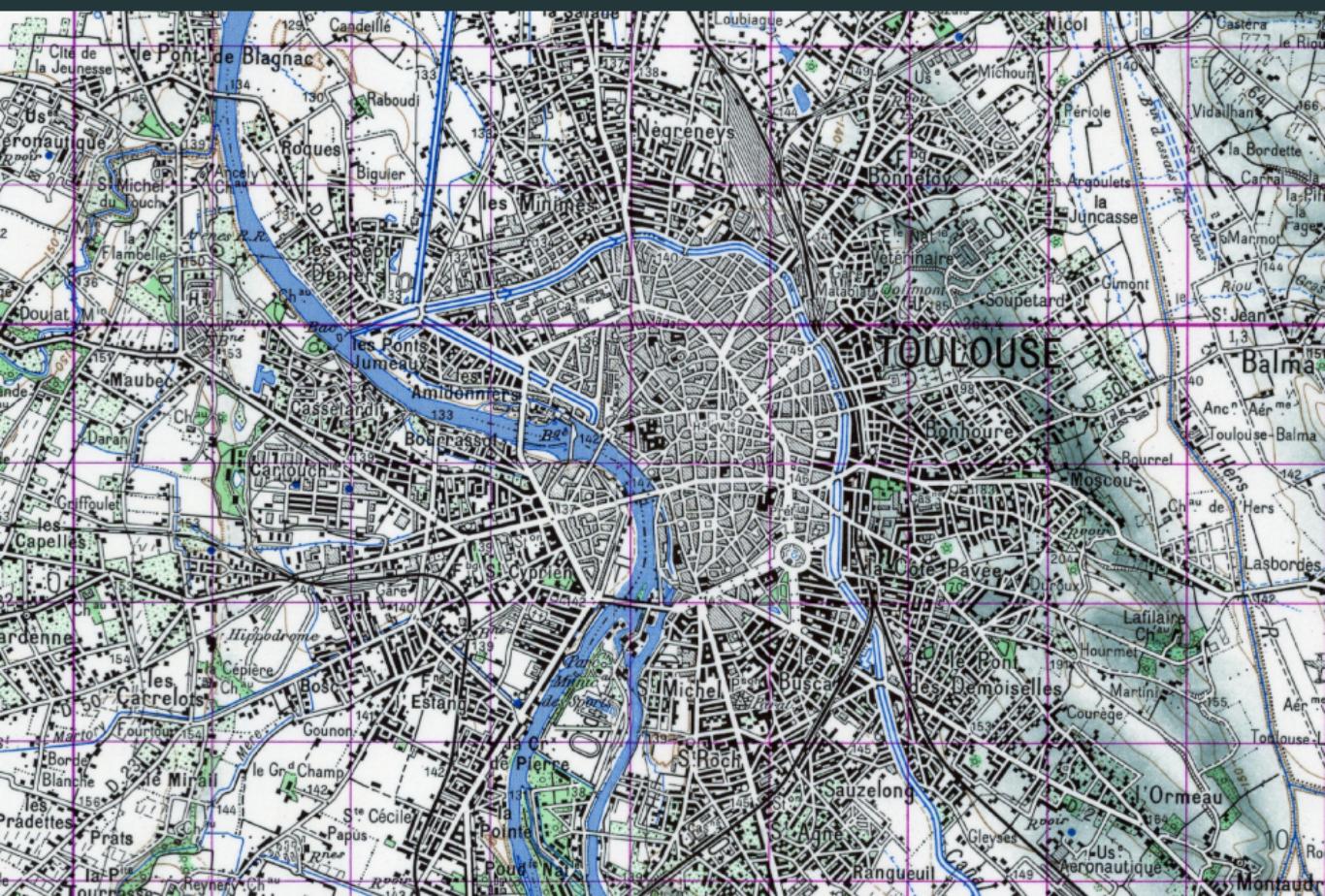


Applications

Données



Applications



Données

Applications

Données

58 Saturday The first part of the day showers of rain & tempest. The Mill and Father Carl Clare & Father Edmund & 34 strong gate. AM sent the 2 Uplanders at Bay. Several Indians at the house. Self and Man seriously employed.

59 Sunday Cloudy and much snow. PM 34 strong gate. Spent the day in truce. Some Captain Cabiling and family on the plantation and got very laboring hard.

60 Monday Clare Cleary & Father Edmund Marshall light. One self and man employed digging for bricks at the Slave near the salt factory. Case signed up 4000 of bricks. Stinkhouse and Squash Pots. EM came back. Wilson and two of the hunters gave them 3 kinds of tobacco & drink and sent them to the tent.

61 Employed at necessary jobs. AM 14 Indians of 4th day have 11. This is the last time they will be here.

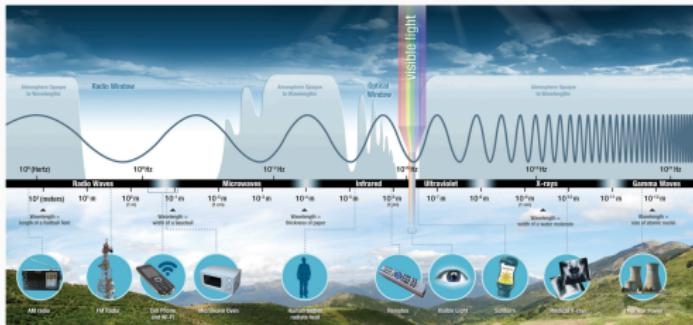
62 Wednesday Weather fine. Leather 18. AM 34 strong gate. The man employed building Martin's self employed at necessary tasks. Martine

63 Thursday AM 20. Father Edmund 18 employed as little. 2 Martins and 2 Cabills is all for Friday.

64 Friday AM 14. Father Edmund 18. Employed at necessary work in and about the house.

65 Saturday The first Mill and Witch rain fine and little employe. AM 34 strong. EM came 2 Indians in. Then ap. and leather 18. brought 4 lots of fish. Bartone boards for the fish and let them have 11. This is the last day. 11 firs to the house to back dinner 22 Oct 1830.

Les images satellites contiennent des bandes¹ spectrales et radar permettant de distinguer matériaux, végétation, etc.



Couverture globale, haute résolution spatiale et temporelle

¹Chaque bande couvre une portion du spectre électromagnétique. On mesure l'absorption, la transmission ou la réflexion du rayonnement.

Une seule prédiction par image, la réponse peut être continue (**régression**) ou discrète (**classification**)



Régression

- Espaces de stationnement
- Densité de végétation
- Pollution atmosphérique

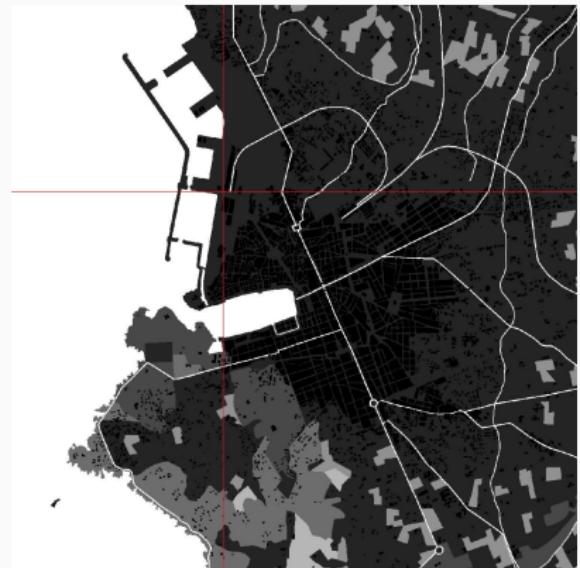
Classification

- État des routes
- Toitures solaires
- Zones de chaleur

La **segmentation** sémantique permet de faire une ou plusieurs prédictions pour chaque pixel de l'image



Image



Prédiction

La **localisation** d'objet vise à localiser (boîte) et à identifier (classe) un nombre arbitraire d'objets dans une image



Image



Prédiction

Il existe de nombreuses **autres applications**, quelques exemples classés par modalité d'entrée et de sortie

Image \Rightarrow Image

- Débruitage d'images
- Super résolution
- Transfert de style

Image \Rightarrow Représentation

- Apprentissage par similarité
- Recherche par contenu
- Détection d'anomalies

Image \Leftrightarrow Texte

- Reconnaissance optique de caractères
- Annotations automatique d'images
- Génération d'images

Approche

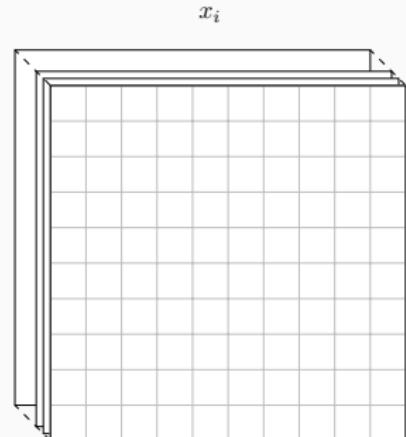
Approche

Une **image digitale** est représentée numériquement comme une matrice multi-dimensionnelle $x_i \in \mathbb{R}^{h \times w \times d}$

Pour chaque pixel...

- Les dimensions h et w contiennent la position spatiale
- Les dimensions d contiennent les caractéristiques ex. couleur RVB

Chaque objet est représenté par des arrangements spatiaux de pixels avec des caractéristiques particulières



$$k = h \times w \times d$$

Approche

Pour chaque application, il existe une fonction inconnue F qui associe chaque image x_i à un label y_i

$$y_i = F(x_i) + \varepsilon_i$$
$$i = 1, \dots, N$$

N est le nombre d'exemples (x_i, y_i) disponibles et ε_i est la partie de y_i qui n'est pas associées à x_i

- La fonction F est très complexe du fait de sa dimensionnalité, des interactions et des non-linéarités
- Approximer ou “apprendre” cette fonction nécessite un modèle flexible et de nombreux exemples (x_i, y_i)

Échantillon labellisé



Approche

On définit une structure de modèle empirique f avec paramètres β pour apprendre cette fonction

$$y_i = f(x_i, \beta) + \epsilon_i$$

On cherche les valeurs minimisant la distance entre la réponse observée et prédite sur un échantillon d'entraînement

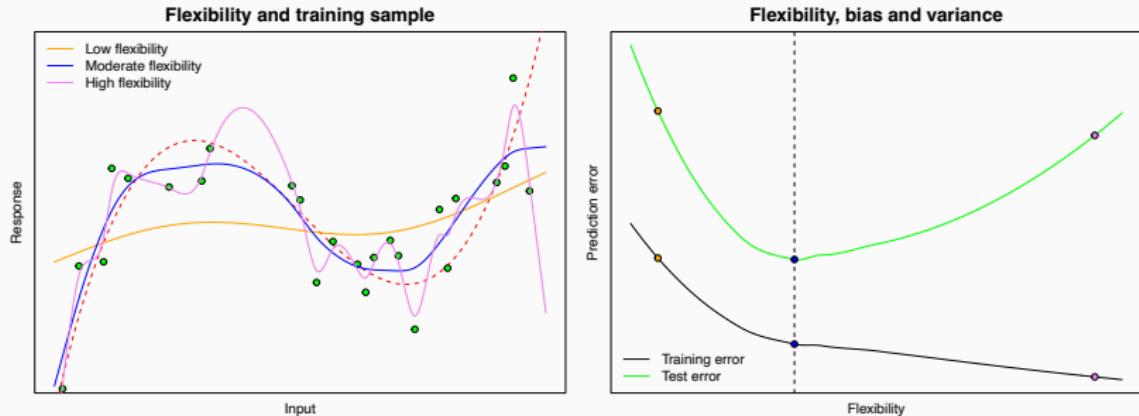
$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \mathcal{L}(y_i, f(x_i, \beta)) + \lambda \rho(\beta)$$

$$i = 1, \dots, n$$

Lorsque $y_i \approx f(x_i, \beta)$ pour de nouvelles observations, le modèle empirique à “apprend” la fonction F

Approche

Les modèles doivent être suffisamment flexibles pour approximer des fonctions complexes, mais ...

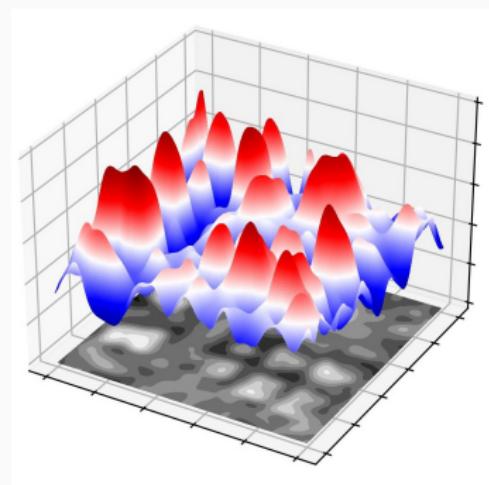


... ils doivent être suffisamment général pour prédire de nouveaux exemples → régularisation, échantillon de test

Modèles

Modèles

Pour “comprendre” une image, le modèle doit extraire de nombreuses caractéristiques localisées de l'image



Ces caractéristiques encodent les relations spatiales (et spectrales) entre les pixels de l'image à plusieurs niveaux

Modèles

Une hiérarchie des caractéristiques allant des interactions les plus locales (simples) au plus globales (abstraites)

Niveau	Caractéristique	Approches
Bas	Contours	Sobel, Canny
↓	Texture	GLCM, LBP
Moyen	Descripteurs locaux	SIFT, ORB, HOG
↓	Descripteurs globaux	Moments, Fourier
Haut	Ex. parties, objets	

Les approches traditionnelles extraient manuellement ces caractéristiques, ingénierie des caractéristiques

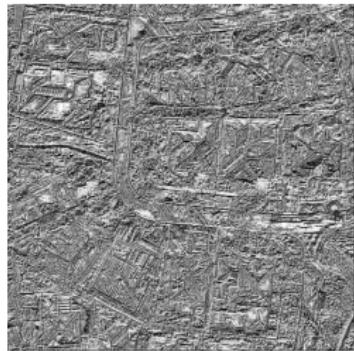
Modèles

Cela revient à faire une hypothèse forte sur la nature du processus générateur de données

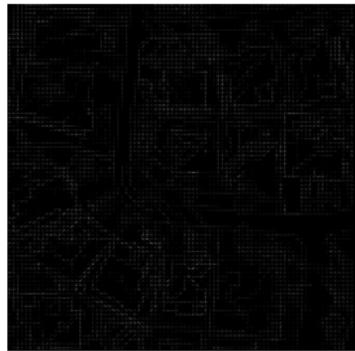
Sobel



LBP

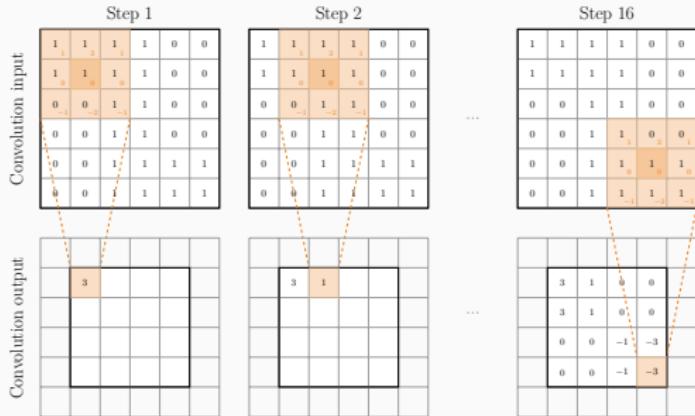


HoG



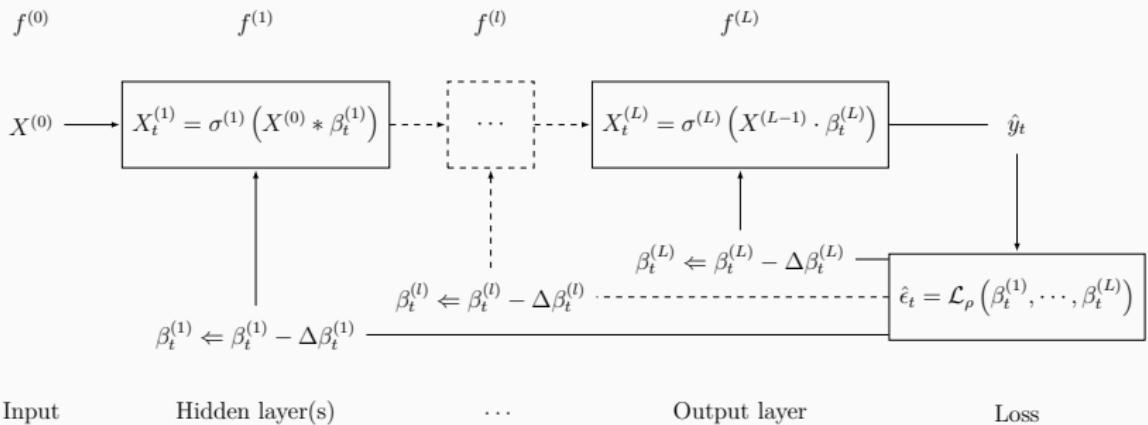
Les couleurs dénotent la présence (blanc) ou l'absence (noir) d'une caractéristique dans une région de l'image

Cette approche peut-être généralisée à l'aide de **convolutions discrètes** paramétrisées donc estimées!



Le noyau se déplace en calculant des sommes d'intensités locales pondérées par les paramètres. Les paramètres d'un noyau correspondent à une caractéristique, les scores de sortie dénotent sa présence ou son absence dans chaque partie de l'image.

Réseau de neurones illustratif



Input

Hidden layer(s)

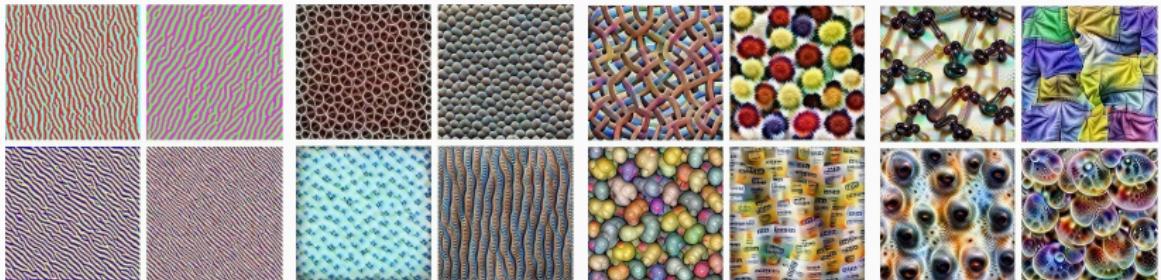
...

Output layer

Loss

L'indice t indique les quantités qui changent à chaque itération de la routine d'optimisation, \Leftarrow est l'opérateur de mise à jour et $\Delta\beta_t^{(l)}$ est la mise à jour des paramètres. La routine d'optimisation met à jour de manière itérative les paramètres pour diminuer l'erreur d'entraînement.

Spécialisation et hiérarchie des caractéristiques



Contours ⇒

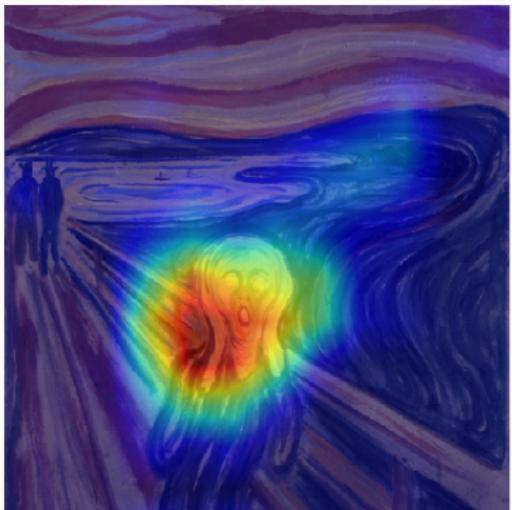
Textures ⇒

Motifs ⇒

Parties

Visualisation des cartes de caractéristiques pour le modèle GoogLeNet (Szegedy et al. 2015) entraîné sur les données ImageNet (Deng et al. 2009). Ces visualisations sont obtenues en fixant les paramètres du modèle et en ajustant les valeurs des pixels d'entrée pour maximiser une représentation spécifique par montée de gradient (Zeiler and Fergus 2014).

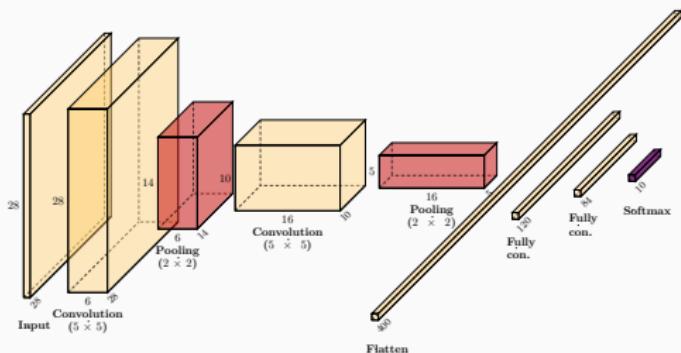
Carte d'attribution



La zone en rouge représente la partie de l'image (Le Cri Edvard Munch, 1893) la plus utilisée par le modèle pour la prédiction (score de peur). Cette carte d'attribution est obtenue en calculant les dérivées partielles de la prédiction vis-à-vis des pixels d'entrée (Selvaraju et al. 2017).

Illustrations

Réseau convolutif LeNet5 (Lecun et al. 1998)



Considérons une seule image. Les structures numériques sont représentées et les connexions entre les neurones sont omises. Par exemple, la première couche convolutionnelle $f^{(1)}$ combine les pixels dans un voisinage de $5 \times 5 \times 1$ en utilisant un ensemble de $d^{(1)} = 6$ noyaux. La couche de *pooling* agrège les valeurs de chaque représentation extraite dans un voisinage de 2×2 .

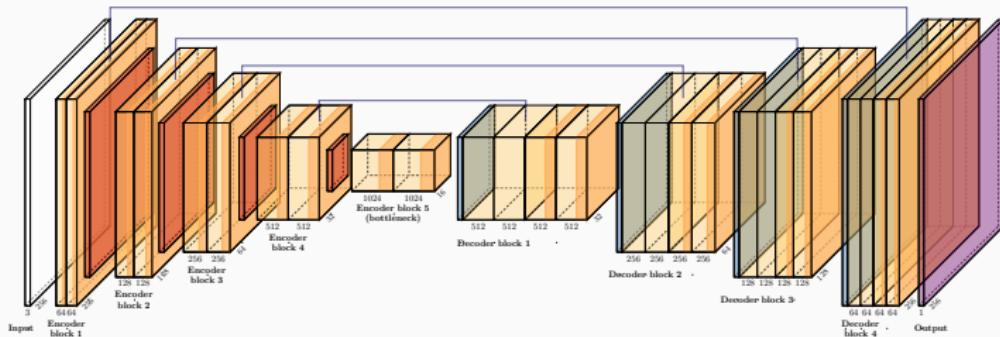
Visualisation d'un réseau convolutif, Harley (2015)

- Étudiez la structure du réseau pour comprendre le fonctionnement des blocs convolutifs

Visualisation des caractéristiques, OpenAI microscope

- Observez comment les représentations sont diverses et organisées de manière hiérarchiques

Réseau convolutif U-Net (Ronneberger et al. 2015)



Les 4 blocs d'encodage contiennent deux couches de convolution avec $d^{(l)} \times 3 \times 3 \times d^{(l-1)}$ paramètres et une activation ReLU, suivis d'une couche de *pooling*. Les 4 blocs de décodage combinent les informations sémantiques (convolution transposée) et les informations spatiales détaillées (connexion résiduelle) à l'aide de deux convolutions. La couche de sortie utilise une convolution ponctuelle avec une activation logistique.

Résumé

Résumé

Disponibilité croissante de données sous forme d'images géolocalisées ex. images satellites ou *street-view*

- Les réseaux de neurones convolutifs permettent de mobiliser efficacement ces données à l'échelle
- Nombreuses applications en gestion urbaine, environnementale, des transports, agricole, etc.
- Facilité de déploiement avec les libraires spécialisées et les modèles fondamentaux pré-entraînés

Merci pour votre attention!

Bibliographie

Bibliographie i

- Lecun, Y. et al. (1998). “**Gradient-based learning applied to document recognition**”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324 (cit. on p. 34).
- Deng, Jia et al. (2009). “**ImageNet: A large-scale hierarchical image database**”. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (cit. on p. 31).
- Belloni, Alexandre et al. (2012). “**Sparse models and methods for optimal instrumentswith an application to eminent domain**”. In: *Econometrica* 80.6, pp. 2369–2429 (cit. on p. 3).

Bibliographie ii

- Natekin, Alexey and Alois Knoll (2013). “**Gradient boosting machines, a tutorial**”. In: *Frontiers in neurorobotics* 7.21 (cit. on p. 4).
- Zeiler, Matthew D. and Rob Fergus (2014). “**Visualizing and understanding convolutional networks**”. In: *Computer Vision - ECCV 2014*, pp. 818–833 (cit. on p. 31).
- Harley, Adam W. (2015). “**An interactive node-link visualization of convolutional neural networks**”. In: *Advances in Visual Computing*. Ed. by George Bebis et al. Springer International Publishing, pp. 867–877 (cit. on p. 35).
- Kleinberg, Jon et al. (2015). “**Prediction policy problems**”. In: *American Economic Review* 105.5, pp. 491–495 (cit. on p. 4).

Bibliographie iii

- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep learning”. In: *Nature* 521, pp. 436–444 (cit. on p. 3).
- Nielsen, Michael A. (2015). **Neural networks and deep learning**. Determination Press.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241 (cit. on p. 36).
- Szegedy, C. et al. (2015). “Going deeper with convolutions”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9 (cit. on p. 31).

- Athey, Susan and Guido W. Imbens (2016). “**Recursive partitioning for heterogeneous causal effects**”. In: *PNAS* 113.27, pp. 7353–7360 (cit. on p. 3).
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.
- Mullainathan, Sendhil and Jann Spiess (2017). “**Machine learning: An applied econometric approach**”. In: *Journal of Economic Perspectives* 31.2, pp. 87–106 (cit. on p. 3).

- Selvaraju, Ramprasaath R. et al. (2017). “**Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization**”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626. DOI: [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74) (cit. on p. 32).

Annexe

Convolution, activation et regroupement

(a) Input image

1	1	1	1	0	0
1	1	1	1	0	0
0	0	1	1	0	0
0	0	1	1	0	0
0	0	1	1	1	1
0	0	1	1	1	1

(b) Convolutional layer
($d^{(1)}$ kernels $3 \times 3 \times d^{(0)}$)

3	1	0	0
3	1	0	0
0	0	-1	-3
0	0	-1	-3

$$h^{(0)} \times w^{(0)} \times d^{(0)}$$

$$h^{(1)} \times w^{(1)} \times d^{(1)}$$

(c) Convolutional layer
(ReLU activation)

3	1	0	0
3	1	0	0
0	0	0	0
0	0	0	0

$$h^{(1)} \times w^{(1)} \times d^{(1)}$$

(d) Pooling layer
(2×2 max.)

3	0
0	0

$$\frac{h^{(1)}}{2} \times \frac{w^{(1)}}{2} \times d^{(1)}$$