Réseaux de neurones IFT 780

Visualisation Par Pierre-Marc Jodoin

Comment visualiser ce qu'un réseau a appris?

VGG16

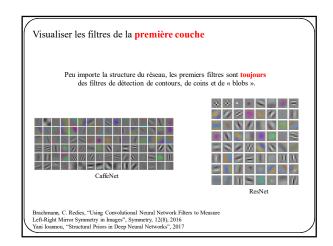


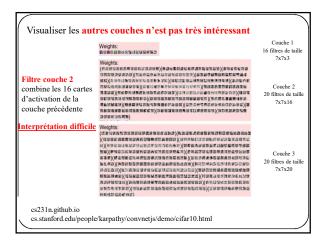


Prédiction 1000 valeurs

3 options:

- 1. Visualiser les filtres
- Visualiser les cartes d'activation (neurones en sortie d'une couche)
 Construire/sélectionner des images qui excitent certains neurones

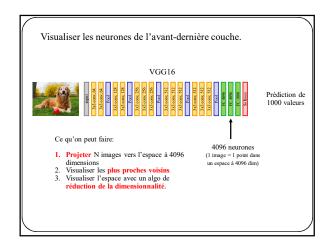




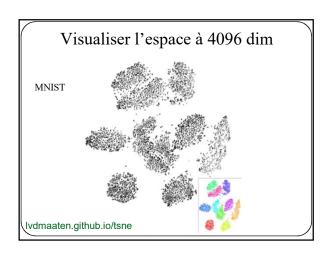
On tire rarement grand-chose à visualiser des filtres. Il faut donc une autre solution:

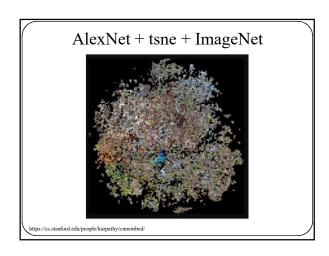
Options restantes:

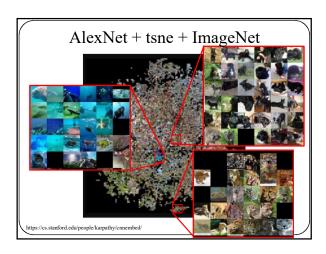
- 1. Visualiser les filtres
- 2. Visualiser les cartes d'activation (neurones en sortie d'une couche)
- 3. Construire/sélectionner une image qui excite certains neurones

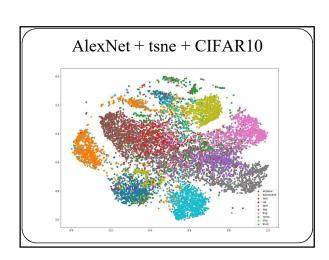






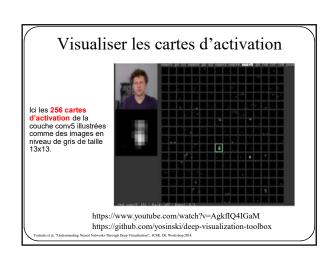








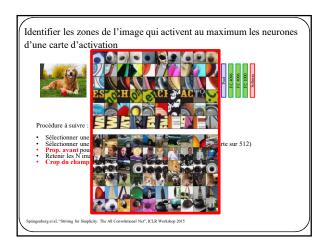


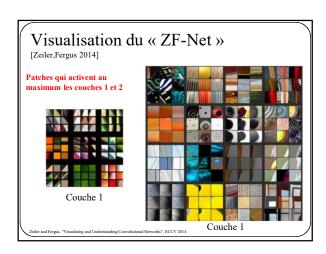


Identifier les zones de l'image qui activent au maximum les neurones d'une carte d'activation

Procédure à suivre :

Sélectionner une couche (ex. conv8)
Sélectionner une carte d'activation de cette couche (ex.: la 20º carte sur 512)
Prop, avant de plusieurs images
Reterir les N images ayant provoqué une activation maximale dans cette carte
Crop du champ récepteur des neurones maximalement activés











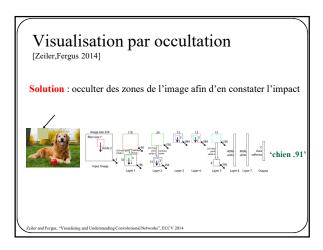
Comment identifier dans cette image ce qui permet au réseau de prédire la classe « chien »?

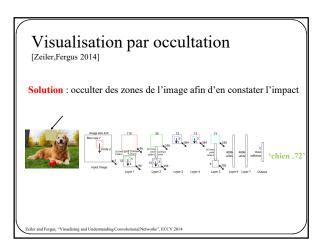


$\begin{tabular}{ll} Visualisation par occultation \\ [Zeiler, Fergus 2014] \end{tabular}$ Solution : occulter des zones de l'image afin d'en constater l'impact

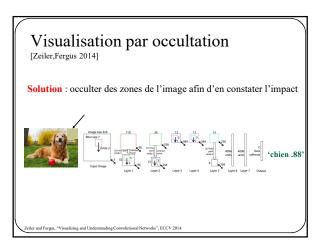


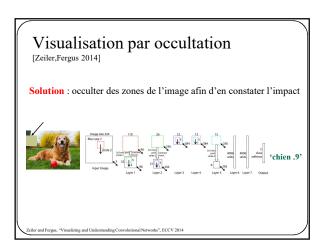
Visualisation par occultation [Zeiler,Fergus 2014] Solution: occulter des zones de l'image afin d'en constater l'impact Solution: occulter des zones de l'image afin d'en constater l'impact Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones d

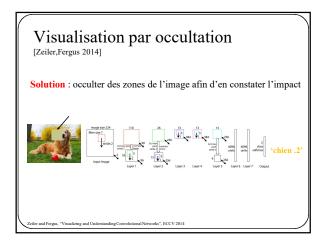


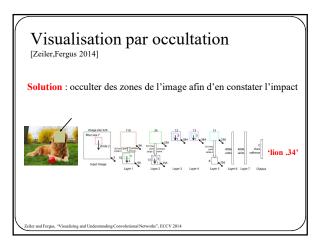


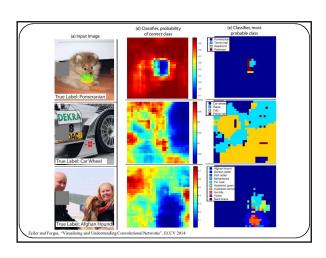
Visualisation par occultation [Zeiler,Fergus 2014] Solution: occulter des zones de l'image afin d'en constater l'impact Solution: occulter des zones de l'image afin d'en constater l'impact Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones de l'image afin d'en constater l'impact Light Solution: occulter des zones d











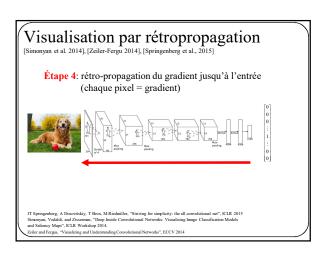
Visualisation par rétropropagation (Simonyan et al. 2014). [Zeiter-Forga 2014]. [Springenbrag et al., 2015] Même question: comment identifier dans cette image ce qui permet au réseau de prédire la classe « chien »? AlexNet Il Springenbra, A Denshida, Tima, Malandia, "Siming de confuire," fond considerant Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education," Chip land Constituent Madels and Education, "Chip land Constituent Madels and Education Madels and Education," Chip land Constituent Madels and Education Ma

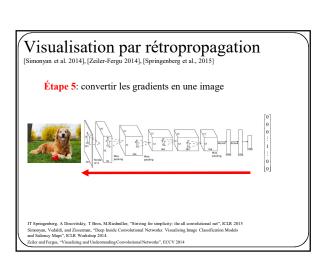
Visualisation par rétropropagation
[Simonyan et al. 2014], [Zeiler-Fergu 2014], [Springenberg et al., 2015]

Étape 2: prop. avant d'une image

Il Springenberg, A Dosovitskiy, T Brox, M Riedmiller, "Striving for simplicity: the all convolutional net", KCLR 2015
Simonyan, Veladid, and Zimeram. "Dops pinds convolutional Network: Visualing Image Classification Models
and Saliency Mary, TCLR Workshop 2014.

Visualisation par rétropropagation [Simonyan et al. 2014], [Zeiler-Fergu 2014], [Springenberg et al., 2015] Étape 3: forcer le score du réseau à 1 pour la class d'intérêt J Springenberg. A Douvitskiy, T Box, M.Riedmille, "Striving for simplicity: the all convolutional net", ICLR 2015 Simonyan, Voladid, and Zaserman. "Desp inside Convolutional Networks: Visualizing large Classification Models and Salency Mayer, ICLR Workshop 2014. Zeller and Vergus. "Visualizing and Understanding Convolutional Networks", ECCV 2014

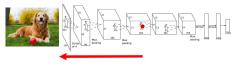




Visualisation par rétropropagation [Simonyan et al. 2014], [Zeiler-Fergu 2014], [Springenberg et al., 2015]

On peut faire la même chose pour 1 neurone :

- Propagation avant jusqu'à la couche X
- Forcer à 0 la sortie de tous les neurones de la couche X
- Mettre à 1 la sortie du neurone d'intérêt
- Propager le gradient vers l'image d'entrée.



Approche par rétropropagation

def generate_grad_saliancy (self, input_image, target_class):

 $model_output = self.model(input_image)$

Init gradients à zero self.model.zero_grad()

 $\label{eq:condition} \begin{tabular}{ll} \# one_hot = 00000100000, 1 sur la classe cible \\ one_hot_output = torch.FioatTensor(1, model_output.size()[-1]).zero_() \\ one_hot_output[0][target_class] = 1 \end{tabular}$

Backward pass model_output.backward(gradient=one_hot_output)

[0] pour éliminer la première dimension (1,3,224,224) gradients_saliancy = self.gradients.data.numpy()[0]

return gradients_saliancy

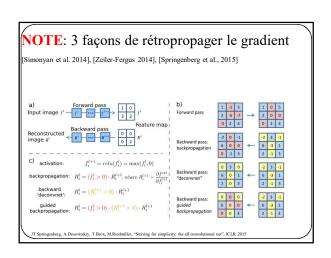
Rétropropagation à partir du score du réseau [Simonyan et al. 2014]

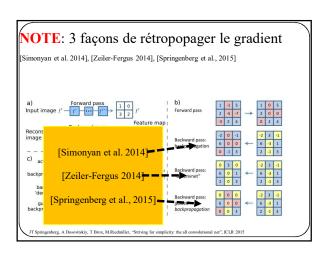


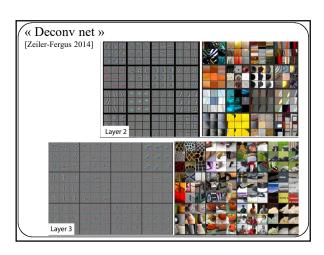


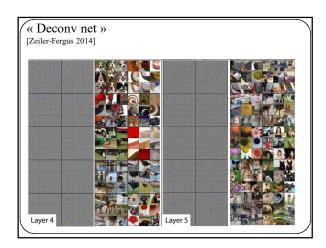


Gradient en valeur absolue

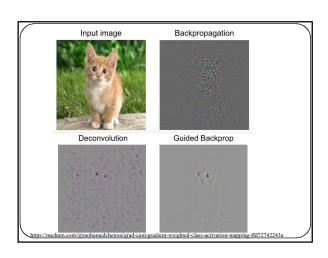




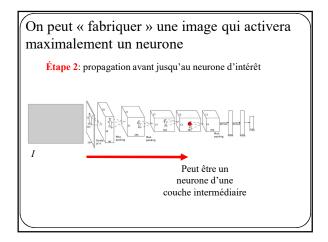


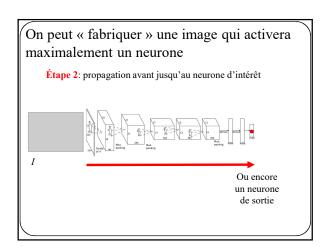


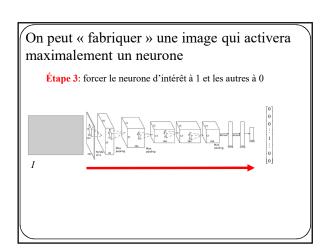
L'approche « *guided backprop* » est plus souvent utilisée car les résultats sont plus saillants et moins bruités

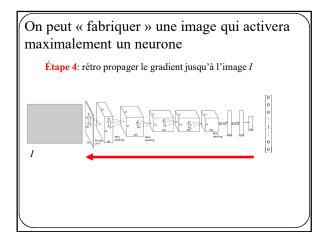


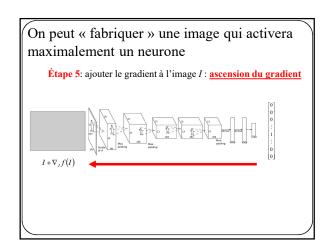
Guided backprop [Springenberg et al., 2015]	
Guided Grad-CAM for "Cat" Guided Grad-CAM for "Dog" https://medium.com/@mohamedchetoui/grad-cam-gradient-weighted-class-activation-mapping-ffd72742243a	
On peut « fabriquer » une image qui activera maximalement un neurone	
Étape 1: Préentraîner le réseau avec une grosse BD (ex. ImageNet)	
The state of the s	
On peut « fabriquer » une image qui activera maximalement un neurone	
Étape 2: Initialiser une image avec des valeurs 0	
I	

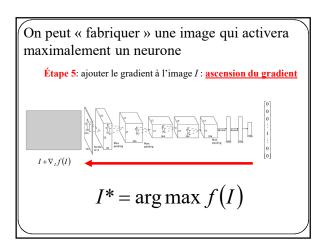












On peut « fabriquer » une image qui activera maximalement un neurone

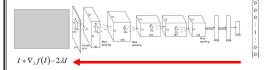
Étape 5: ajouter le gradient à l'image I: ascension du gradient

Afin de s'assurer que l'image produite soit lisse, on rajoute un terme de régularisation, souvent de type L2

$$I^* = \arg\max f(I) - \lambda ||I||^2$$

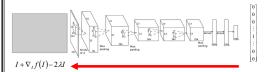
On peut « fabriquer » une image qui activera maximalement un neurone

Étape 5: ajouter le gradient à l'image I: ascension du gradient



$$I^* = \arg\max f(I) - \lambda ||I||^2$$

On peut « fabriquer » une image qui activera maximalement un neurone



NOTE IMPORTANTE: pour cette opération, les poids du réseau sont gelés seule l'image d'entrée est modifiée

