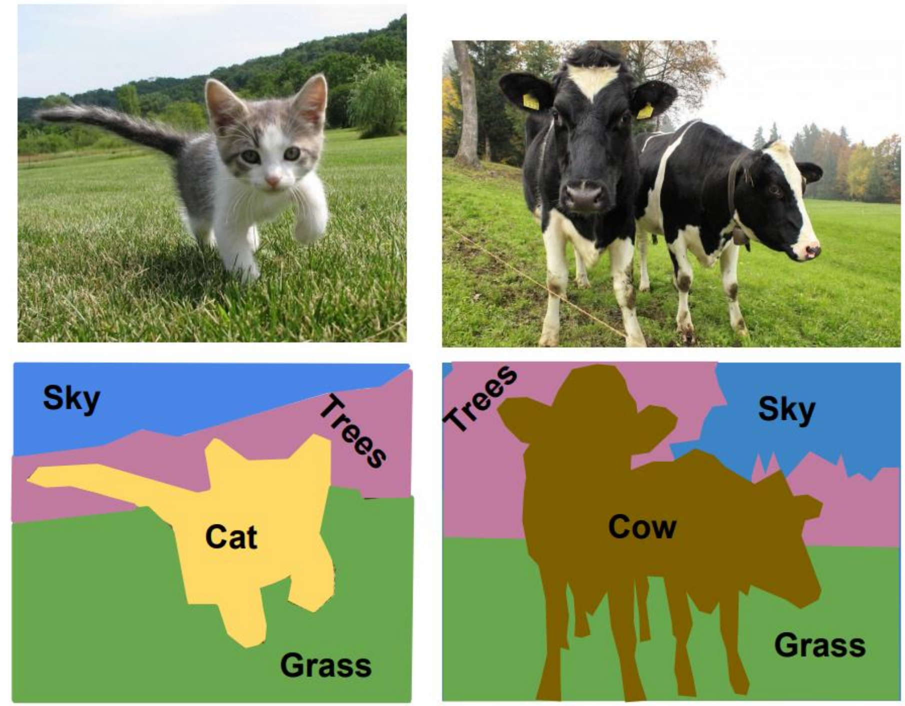
Segmentation

Segmentation sémantique

- Attribuer une classe pour chaque pixel
- Pas tenir comte des différentes instances
  - toutes les vaches, arbres, etc. sont groupés ensembles



- Limite les applications : ne va pas distinguer les voitures, empêchant le *tracking*

Adapté de cs231n

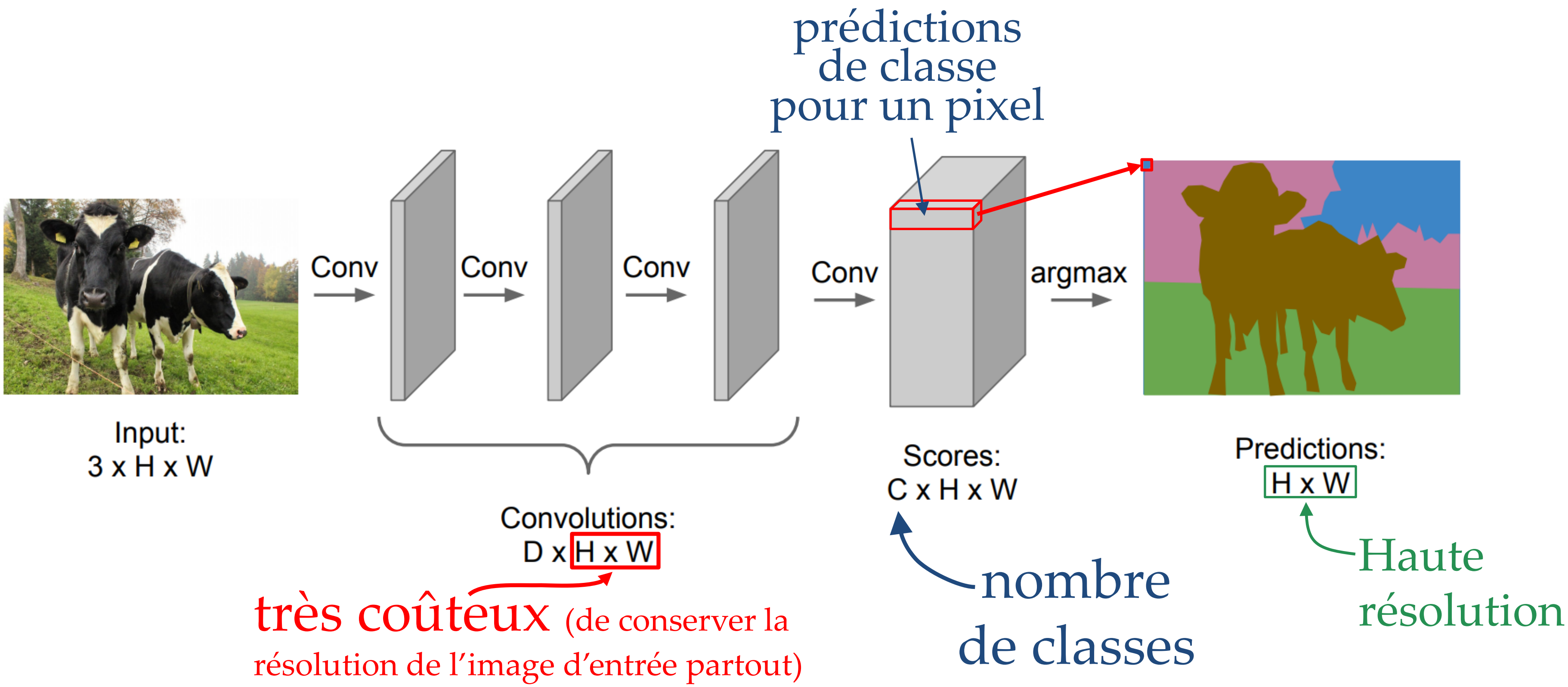
Segmentation sémantique

- S'entraîne en calculant la perte pour chacun des pixels
  - nécessite d'avoir la vérité-terrain pour **TOUS LES PIXELS** ($$$)
- Souvent, fort débalancement des classes (**`background`** dominant)
  - augmenter la perte pour les classes moins représentées
  - *focal loss* peut aider à réduire l'impact

Focal Loss for Dense Object Detection, Lin et al., ICCV 2017

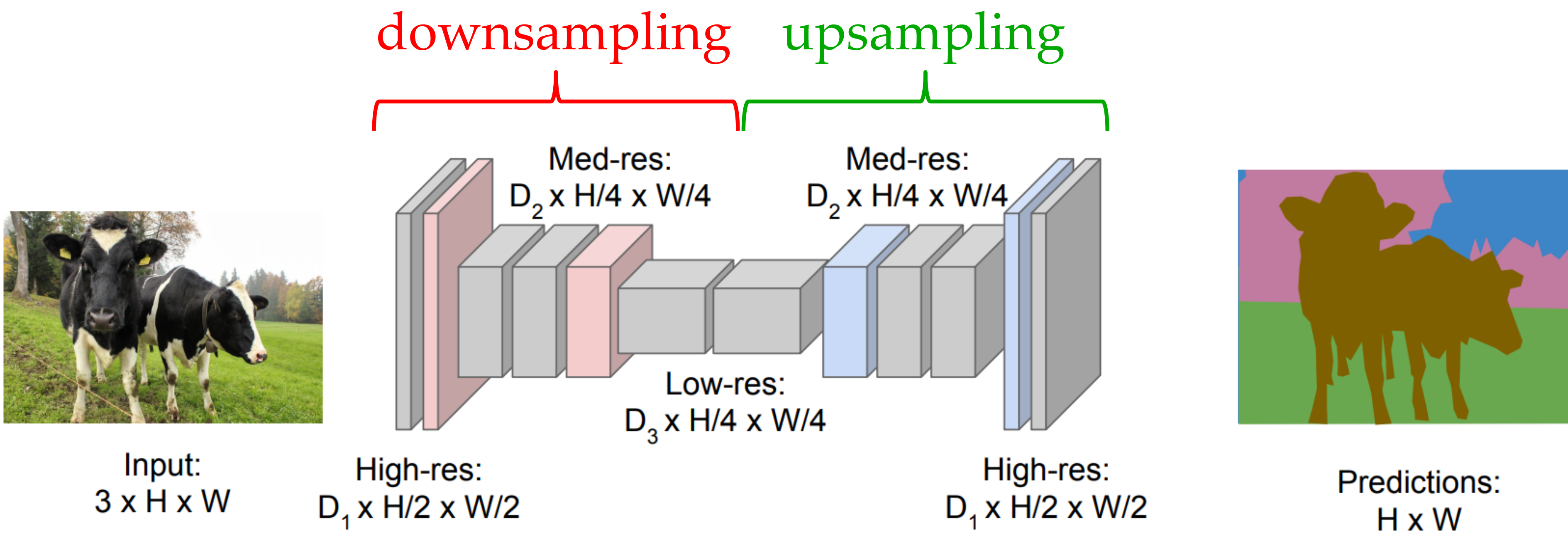
(naïve) Approche Fully-Convolutional

- Dernières couches de convolutions agissent comme des *fully-connected*, pour faire la classification



Adapté de cs231n

Hourglass : réduire les calculs

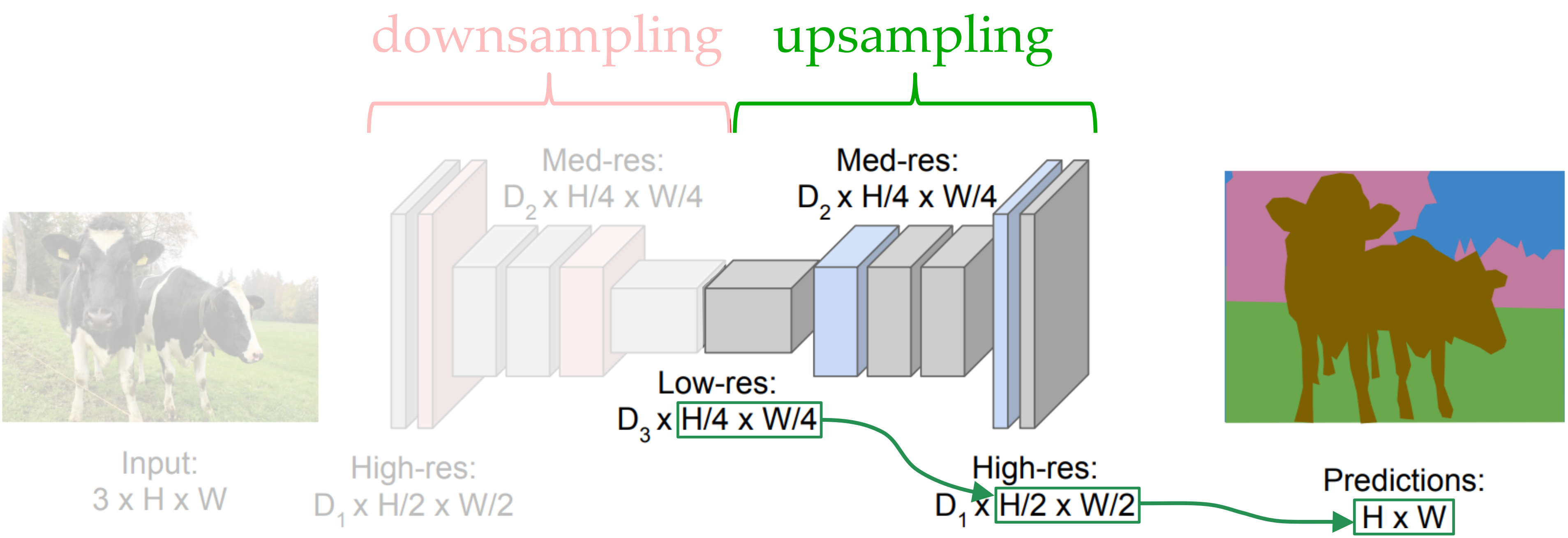


- Perte d'information spatiale fine
  - baisse de résolution H, W
  - *verra plus loin comment corriger cela*

Adapté de cs231n

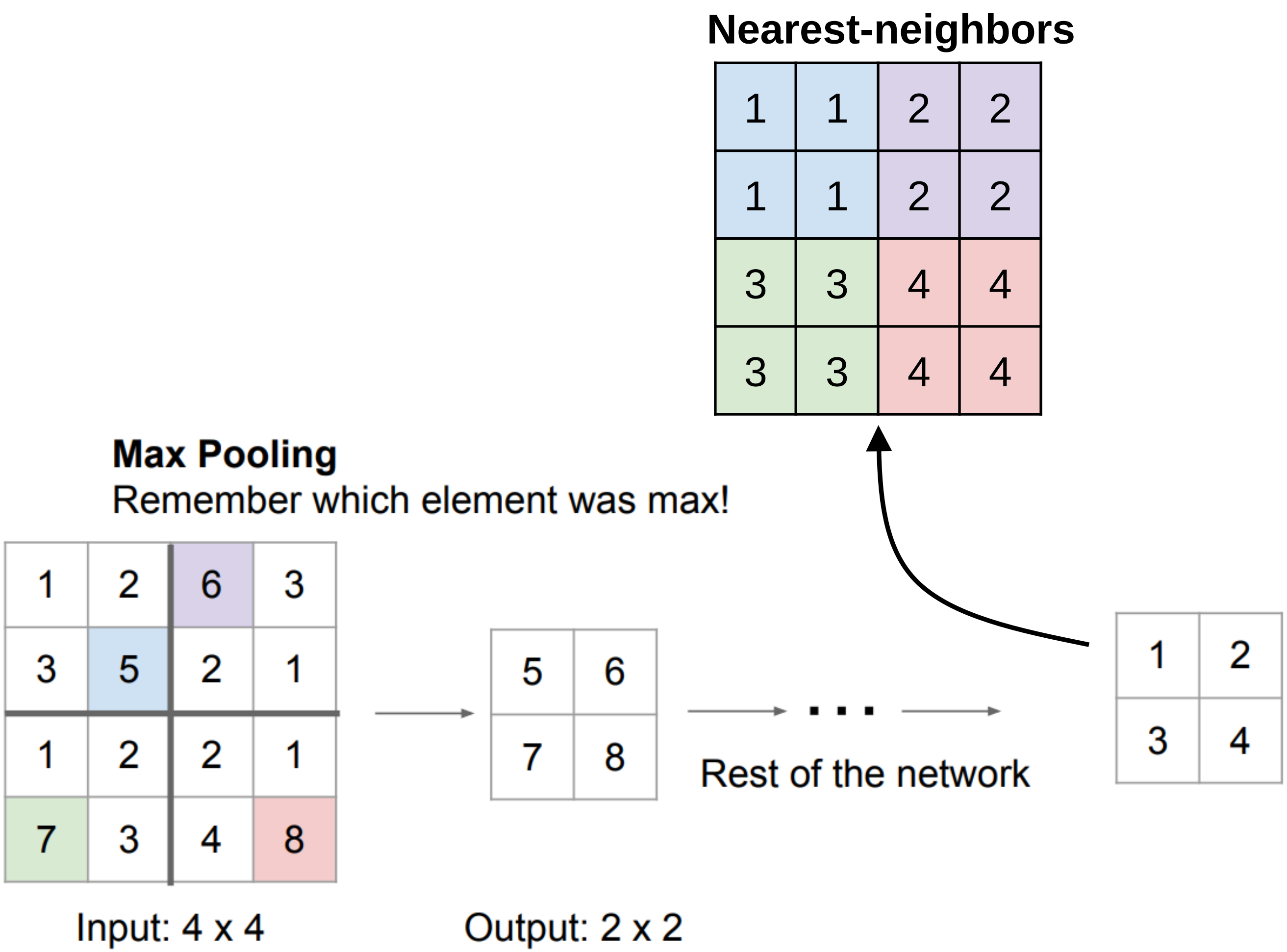
Opération d'upsampling

- Pour augmenter la résolution spatiale
- Plusieurs approches possibles



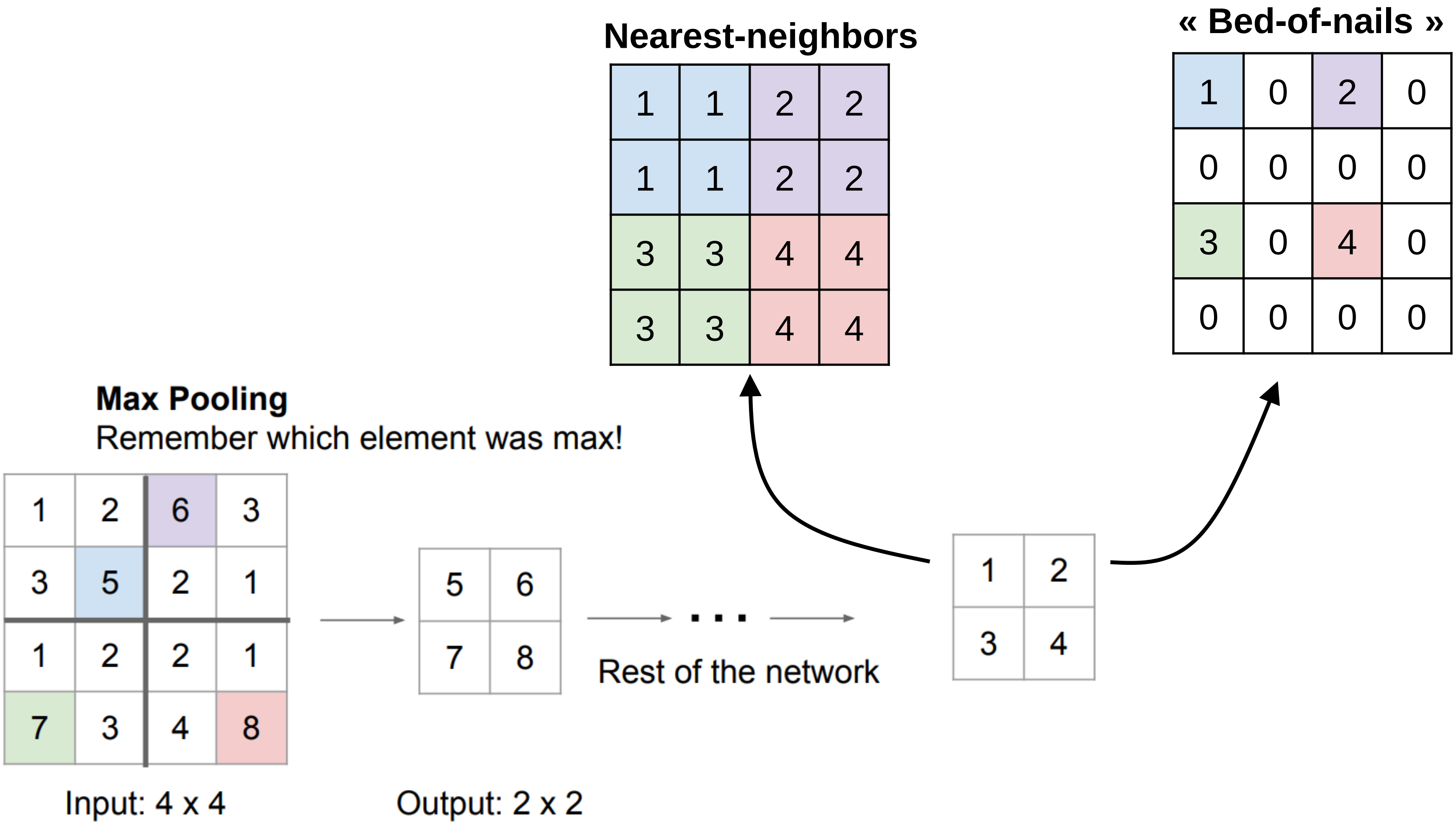
Adapté de cs231n

Upsampling : sans paramètres

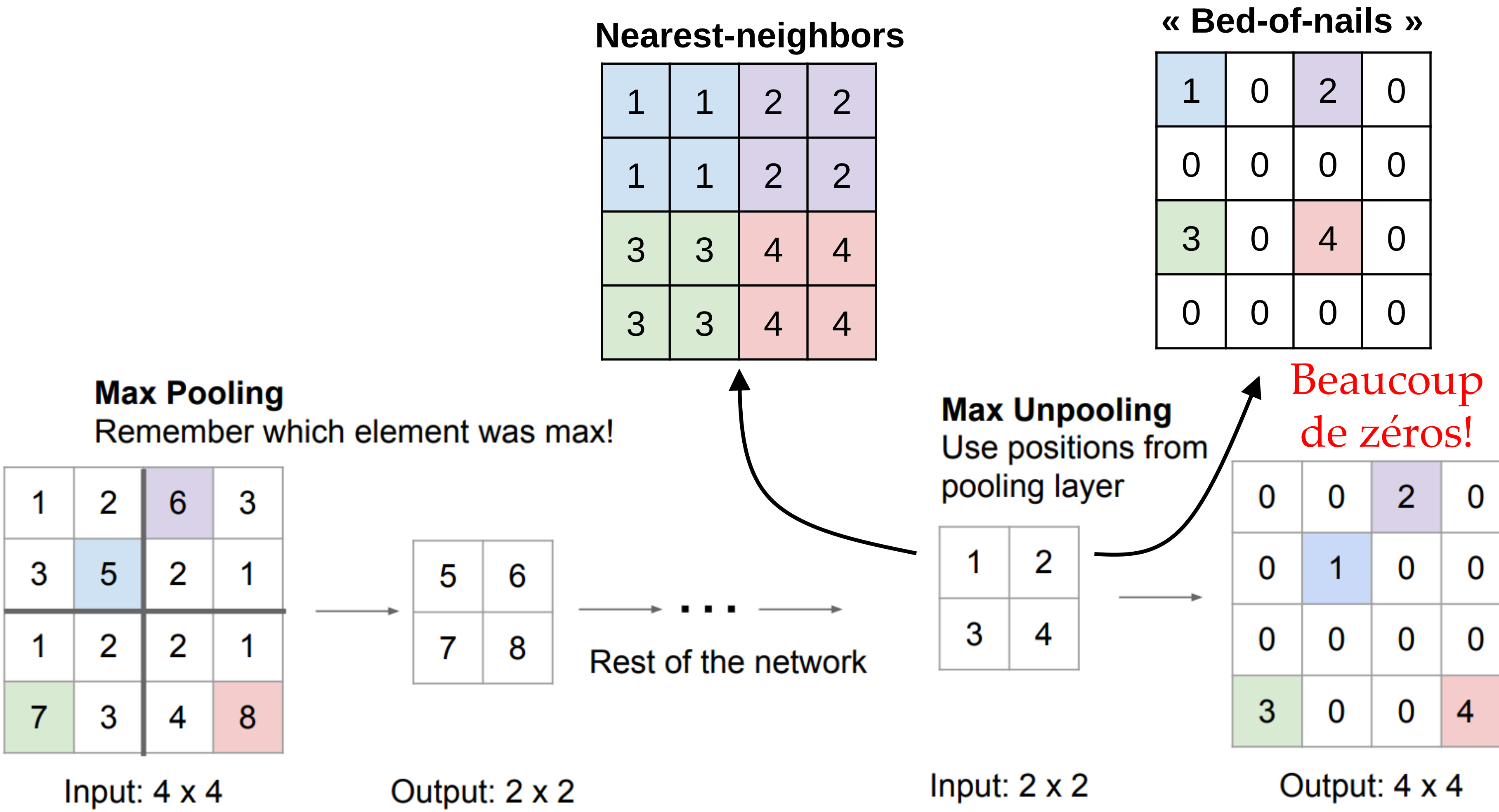


Adapté de cs231n

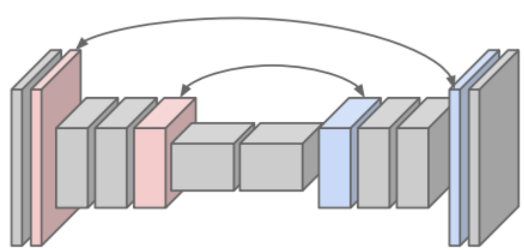
Upsampling : sans paramètres



Upsampling : sans paramètres

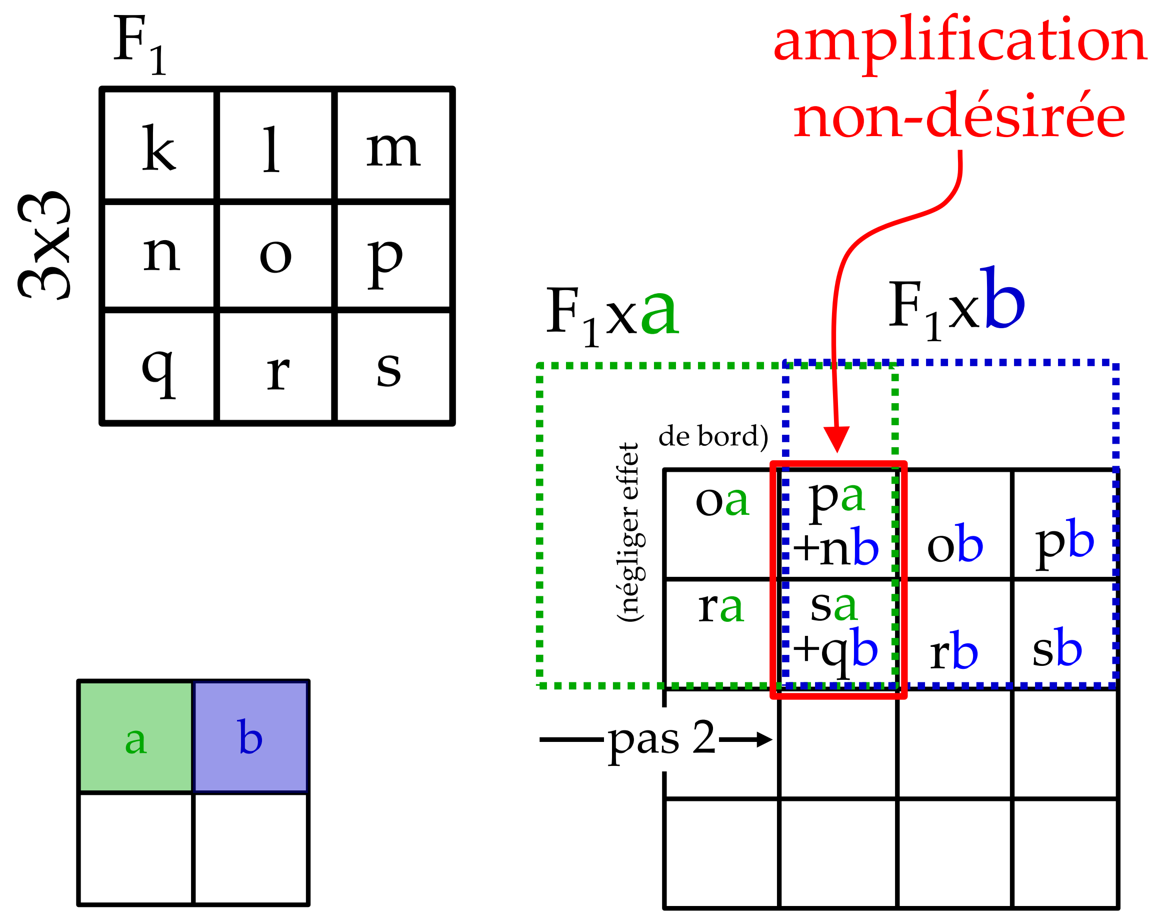


Corresponding pairs of downsampling and upsampling layers



Upsampling avec paramètres : upconvolution

- Des filtres *F* appris
- Exemple : filtre 3x3, pas de 2
- Peut créer des artéfacts en forme d'échiquier dans la sortie
- S'évite avec
  - filtre 2x2 stride 2
  - filtre 4x4 stride 2
  - etc…



Upsampling 2x2 ⇒ 4x4

Other names:
-Deconvolution (bad)
-Upconvolution
-Fractionally strided convolution
-Backward strided convolution

Adapté de cs231n

Upsampling : sub-pix + shuffle

Checkerboard artifact free sub-pixel convolution

A note on sub-pixel convolution, resize convolution and convolution resize

Andrew Aitken*, Christian Ledig*, Lucas Theis*, Jose Caballero, Zehan Wang, Wenzhe Shi*
Twitter, Inc.[1]

entrelacement

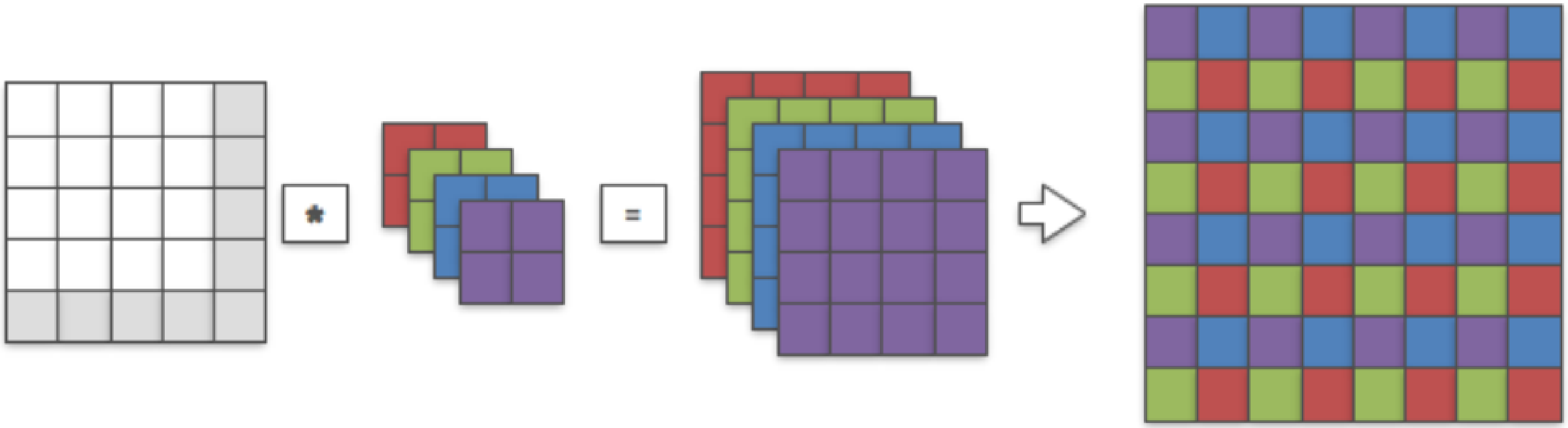
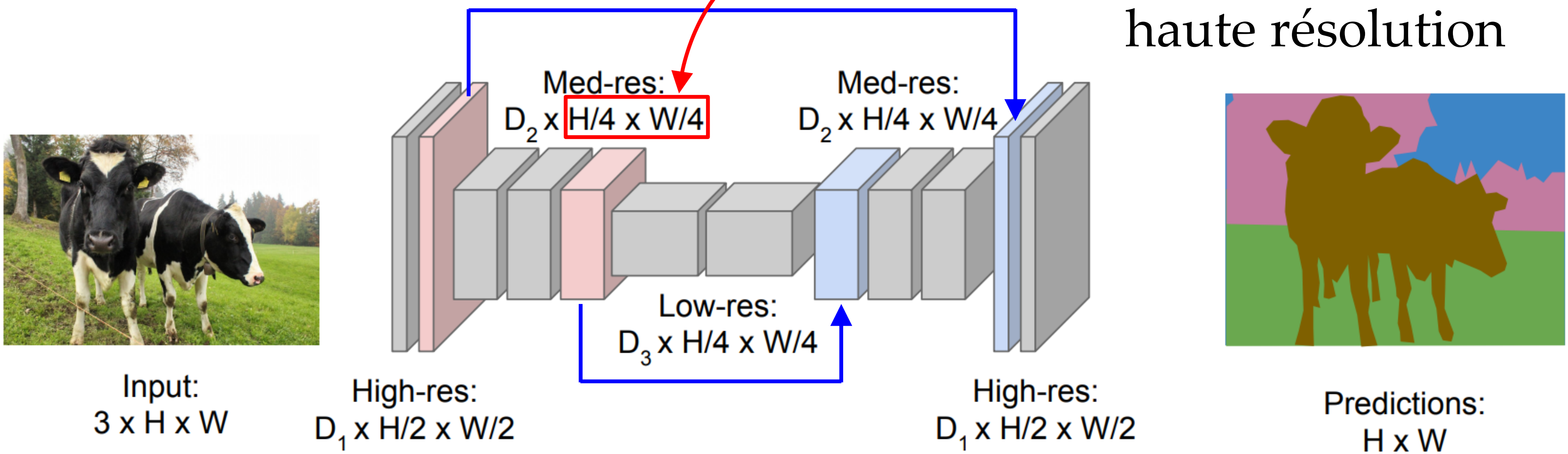


Figure 2: Sub-pixel convolution can be interpreted as convolution + shuffling.

Perte d'information spatiale fine

- Approche *hourglass* fait **perdre l'information** de haute résolution

… nécessaire pour segmentation à haute résolution



- Pour compenser, ajouter des ***skips connections*** entre les couches de même résolution : U-Net
- Rappel : ***skip*** n'ajoute pas de paramètres

Adapté de cs231n

U-net

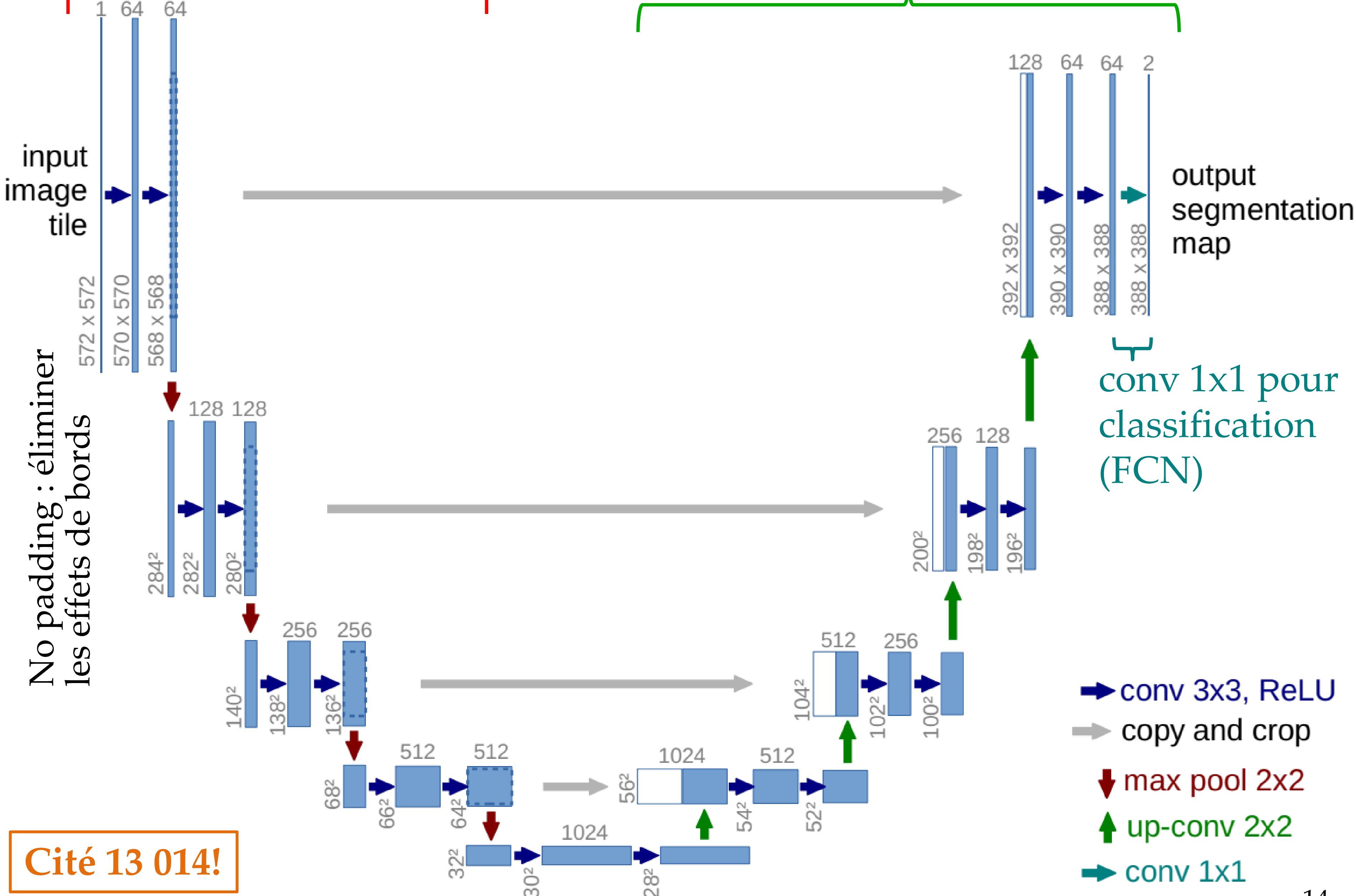
- Vient du domaine de l'imagerie biomédicale
- Quantité de données plus limitée
- Emploie des déformations élastiques pour faire *data augmentation*
  - rappel : comprendre la physique du problème
  - tissus humains sont mous et déformables
- Perte qui tient compte de la position
  - perte plus élevée pour les erreurs sur background en pourtour des autres classes

U-Net: Convolutional Networks for Biomedical Image Segmentation, Ronneberger et al., MICCAI 2015

U-net

downsampling

upsampling



No padding : éliminer les effets de bords

conv 1x1 pour classification (FCN)

Cité 13 014!

U-Net: Convolutional Networks for Biomedical Image Segmentation, Ronneberger et al., MICCAI 2015

Segmentation d’instances

Segmentation d'instances

- Hybride entre **segmentation** et **détection**
- Peut le percevoir comme le problème complet de la vision 2D

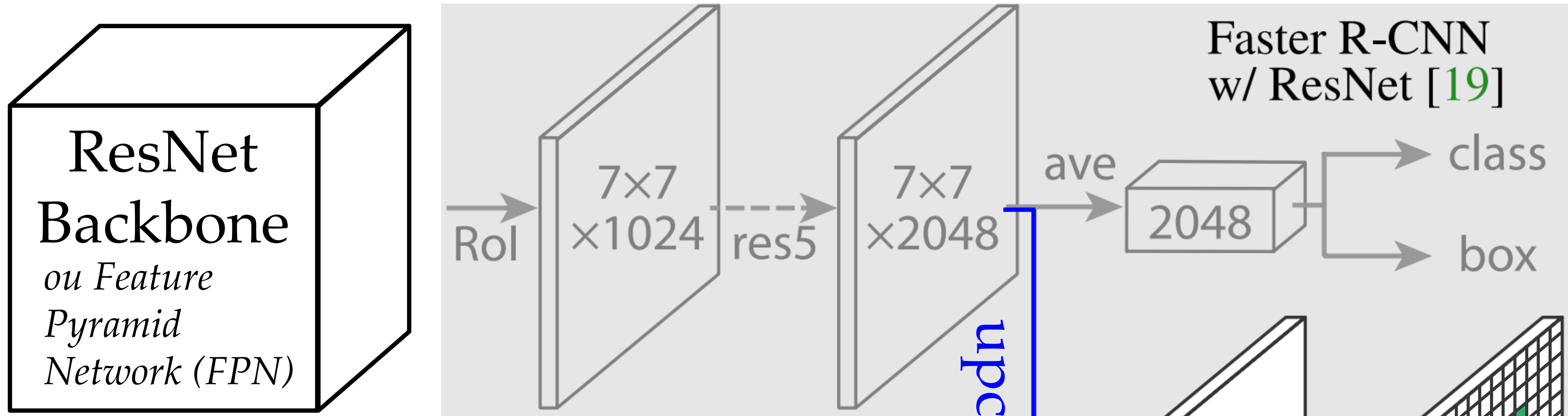
Tiré de cs231n

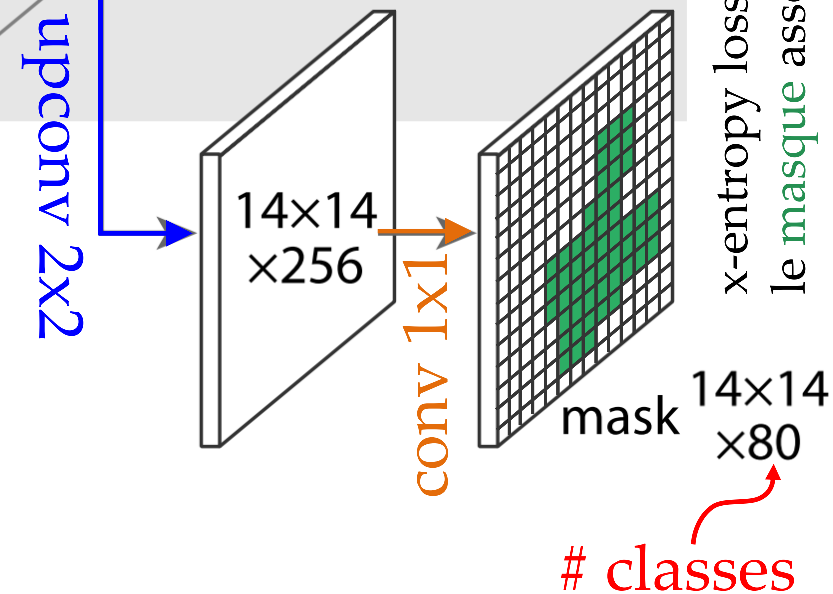
DOG, DOG, CAT

Mask R-CNN

- Performe de la segmentation d’instances
- Modification de Faster R-CNN
  - Simple ajout d’une tête supplémentaire pour prédire un masque binaire !
- N’ajoute qu’un tout petit surplus de calcul
- RoIPool → RoIAlign
- 200 ms par trame (*frame*)

Mask R-CNN : Modification à Faster R-CNN



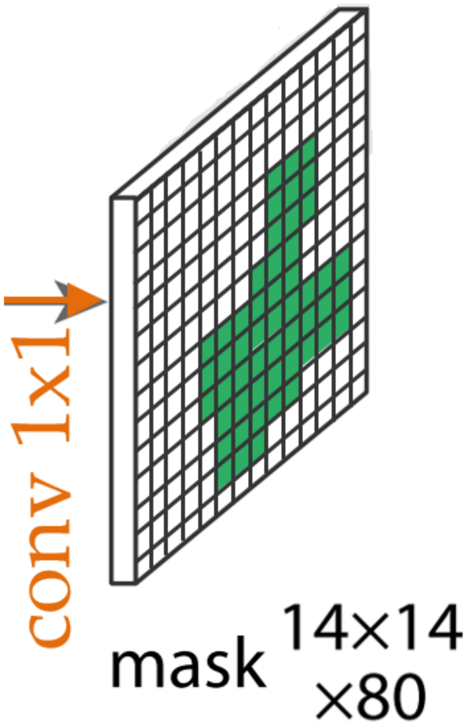


x-entropy loss par pixel, seulement pour le masque associé à la bonne classe

- upconv pour augmenter résolution
- FCN 1x1 comme U-Net, pour prédire les masques
- Découplage :
  - un masque par classe
  - élément clé des performances

Fully convolutional instance-aware semantic segmentation, CVPR 2017.

Découplage des masques



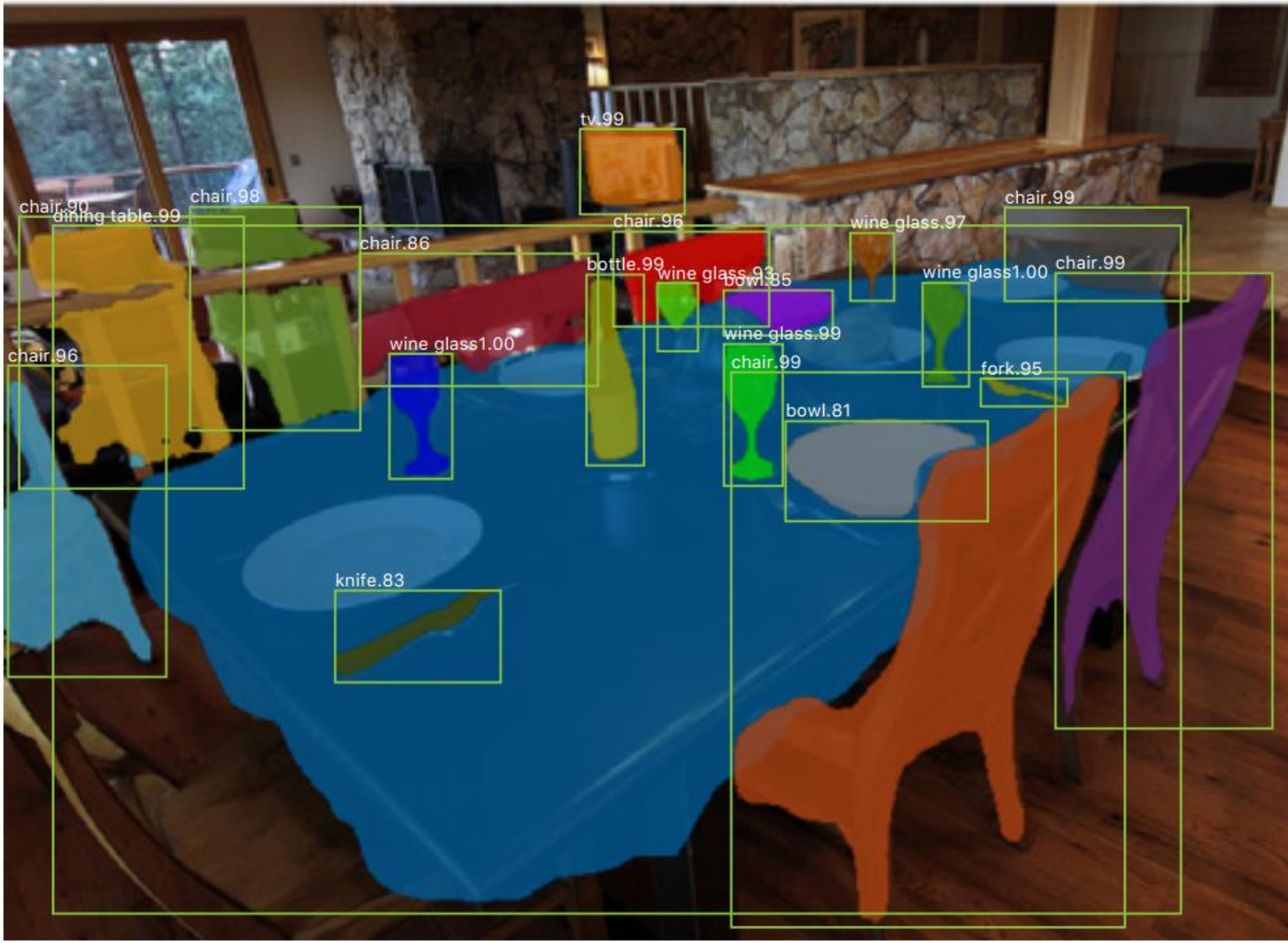
$$softmax : \; \hat{y}_i = \frac{\exp(z_i)}{\left.\sum_{j \in groupe} \exp(z_j)\right\}\text{couplage}}$$

	AP	AP_{50}	AP_{75}
softmax	24.8	44.1	25.1
sigmoid	**30.3**	**51.2**	**31.5**
	+5.5	*+7.1*	*+6.4*

(b) **Multinomial *vs.* Independent Masks** (ResNet-50-C4): *Decoupling* via per-class binary masks (sigmoid) gives large gains over multinomial masks (softmax).

AP : average precision

Mask R-CNN : exemple



Exemple de segmentation

- Données de *Cityscape*

Conclusion

- Réseaux de neurones s'appliquent aussi sur des problèmes de vision numérique au-delà de la classification
- **Détection**
  - doit retrouver les objets d'une même catégorie, sans connaitre au préalable leur nombre
  - méthodes basés sur propositions (300-2000)
  - concept d'anchor box
  - les plus rapides : proposition par réseau (RPN)
  - régression des anchor box pour localisation
  - existe aussi des méthodes par grille (YOLO)

Conclusion

- **Segmentation sémantique**
  - cherche à trouver la classe de chaque pixel
  - architecture U-Net
    - hourglass : downsampling et upsampling
    - méthodes d'upsampling (upconv)
    - skip connexion pour préserver l'info. spatiale fine
- **Segmentation d'instance**
  - hybride entre détection + segmentation
  - Mask R-CNN : ajout d'une branche *masque*
    - importance de découpler les masques entre les classes
- Fully-convolutional network (FCN)