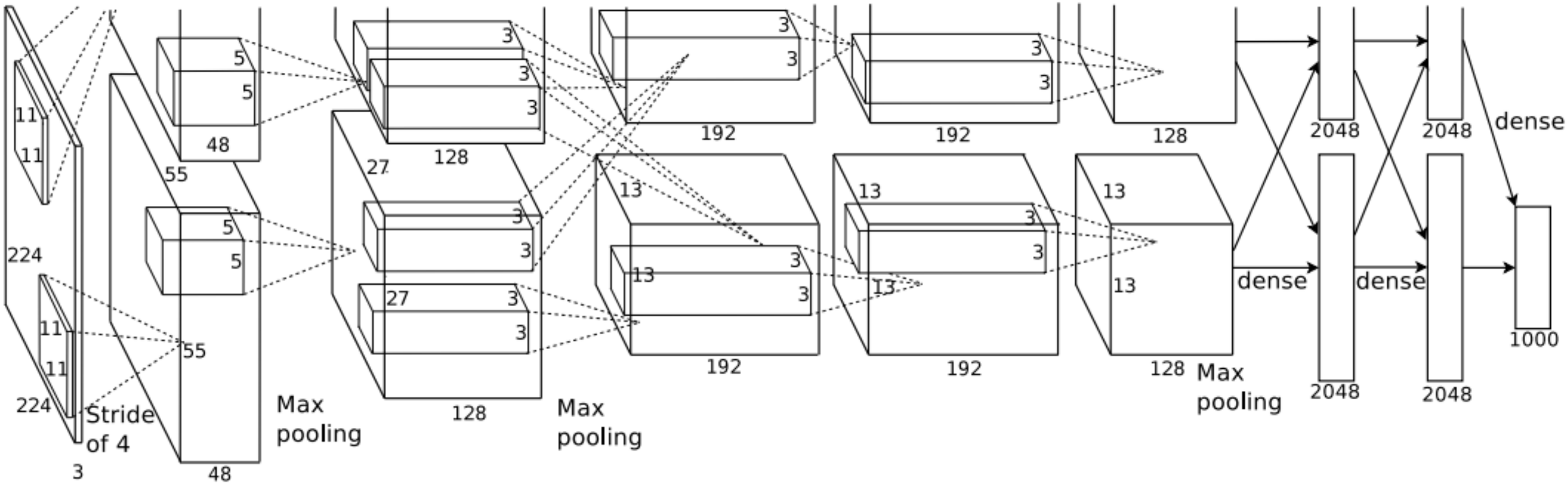
Génération d'images par apprentissage profond



GIF-4105/7105 Photographie Algorithmique, Hiver 2018

Jean-François Lalonde

Merci à Alexei Efros, James Hays, Philip Isola, Andrew Owens, Andrea Vedaldi, Derek Hoiem

facebook

140 milliard d’images
6 milliard ajoutée à chaque mois

flickr

6 milliard d’mages

the simple image sharer
imgur

1 milliard d’images
accédées par jour

YouTube

72 heures téléversées à
chaque minute

3.5 trillion photographs

90% du traffic sur Internet sera des données *visuelles*

30% des vidéos sur Youtube ont moins de 10 vues

«Digital Dark Matter»

[Perona 2010]

Défi principal

- Comment utiliser toutes ces données?
- Avec l'apprentissage profond!
- Mais tout d'abord, utilisons un exemple de système (sans «deep learning») qui utilise des données massives

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

Diffusion

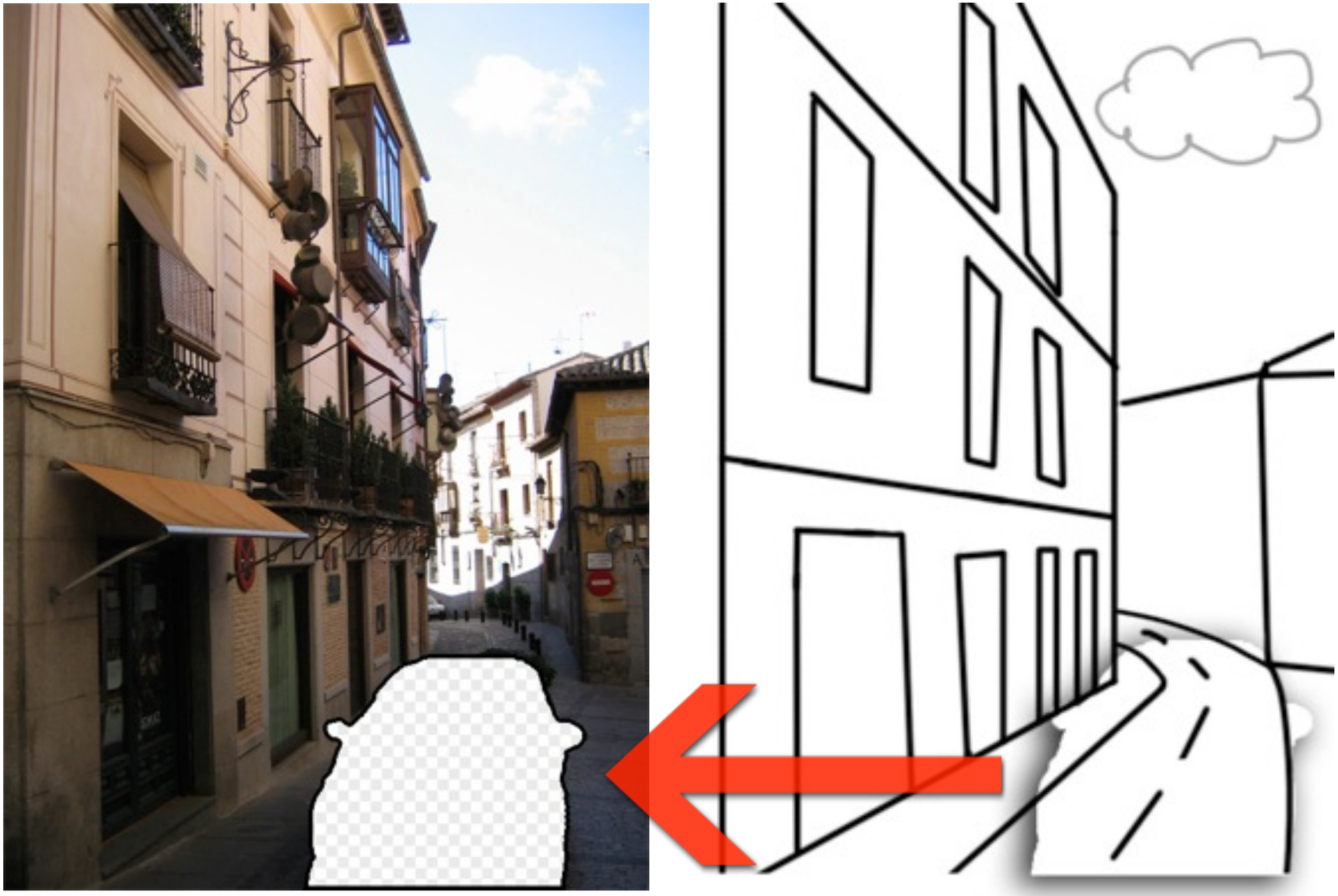
J. Hays

Efros and Leung

J. Hays

J. Hays

Compléter l’image par appariement de scènes

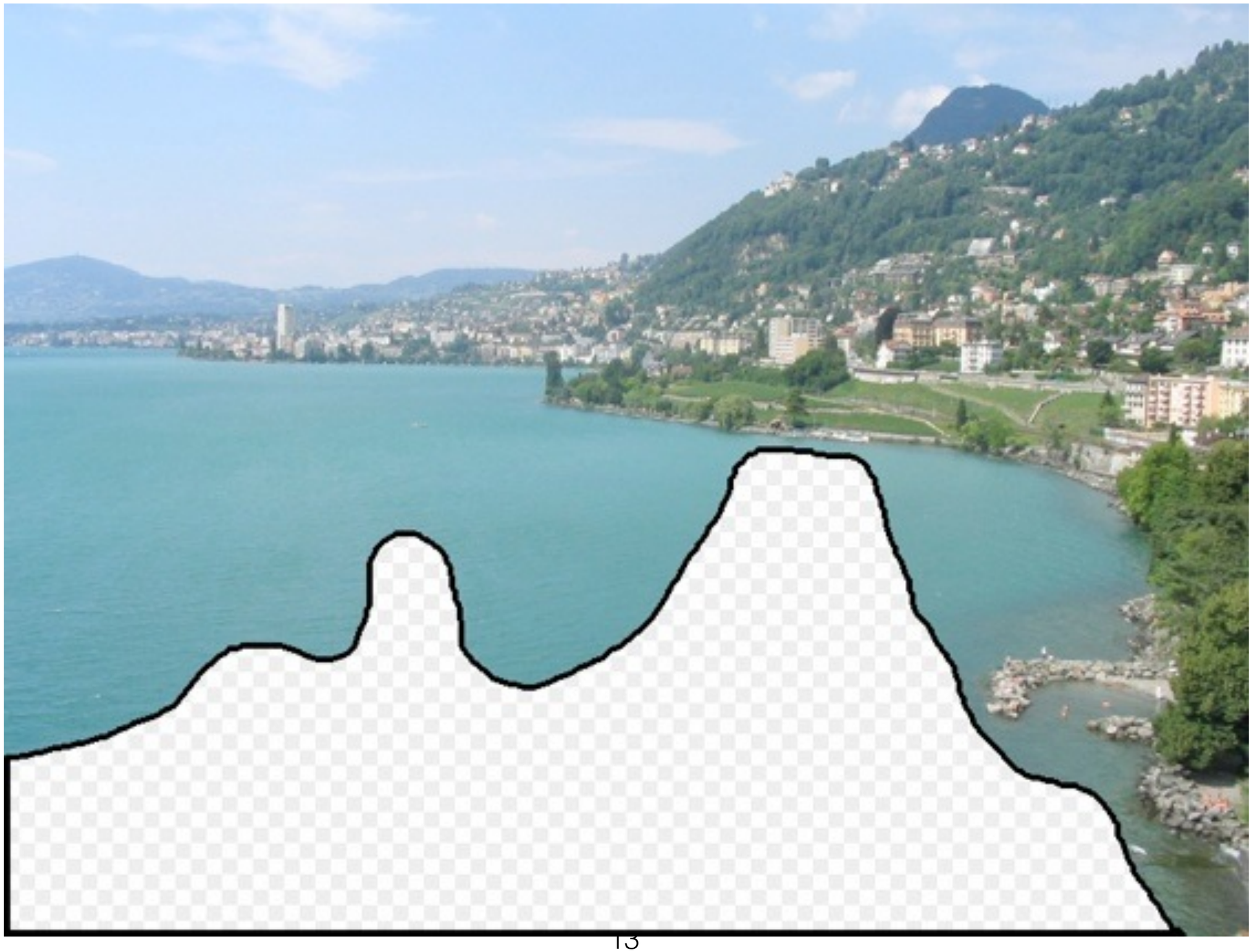


J. Hays

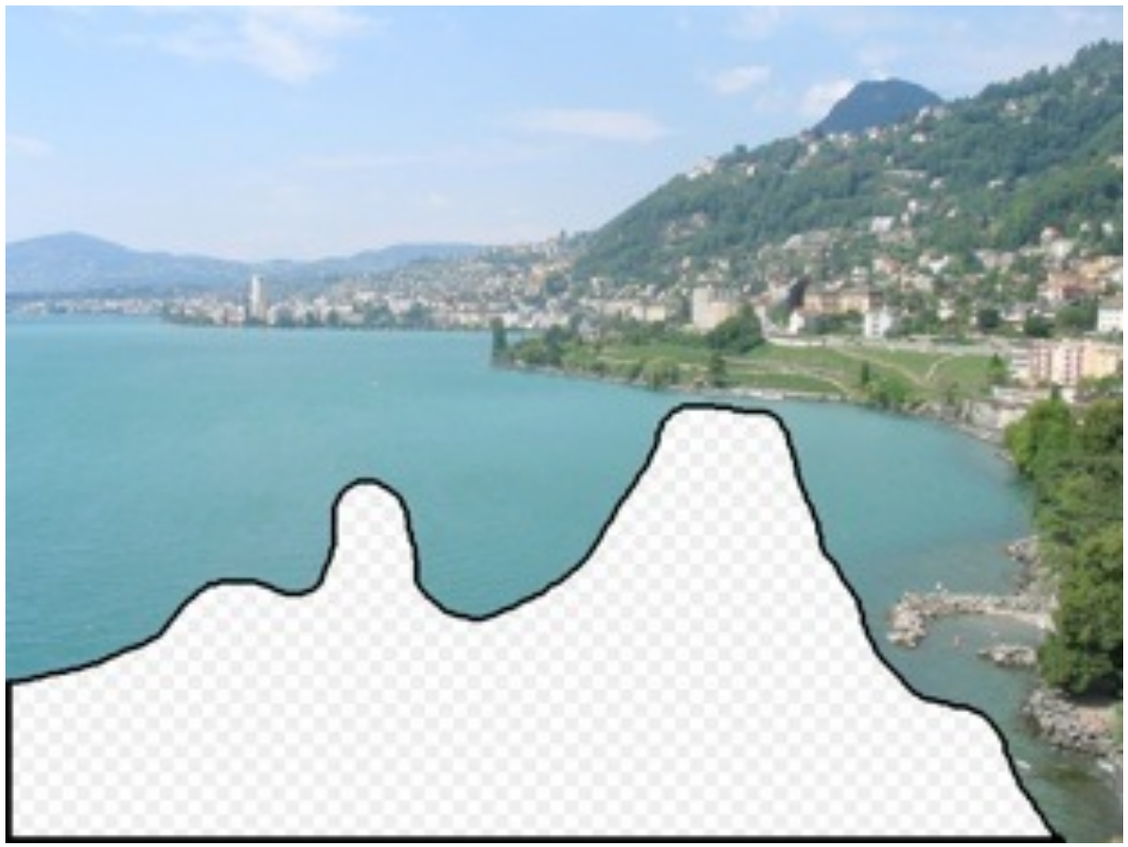
“Scene Completion”

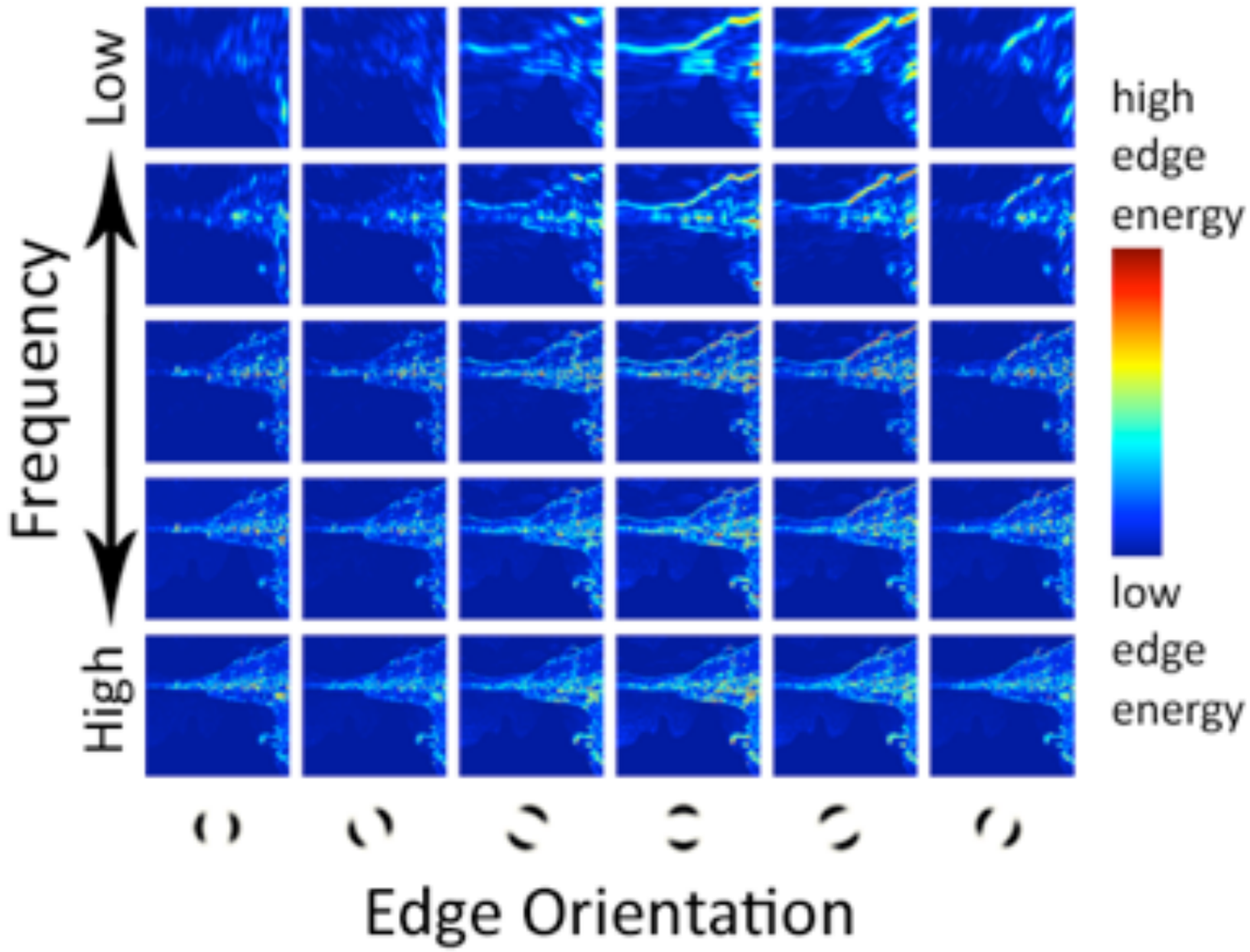
Algorithme

Appariement de scènes



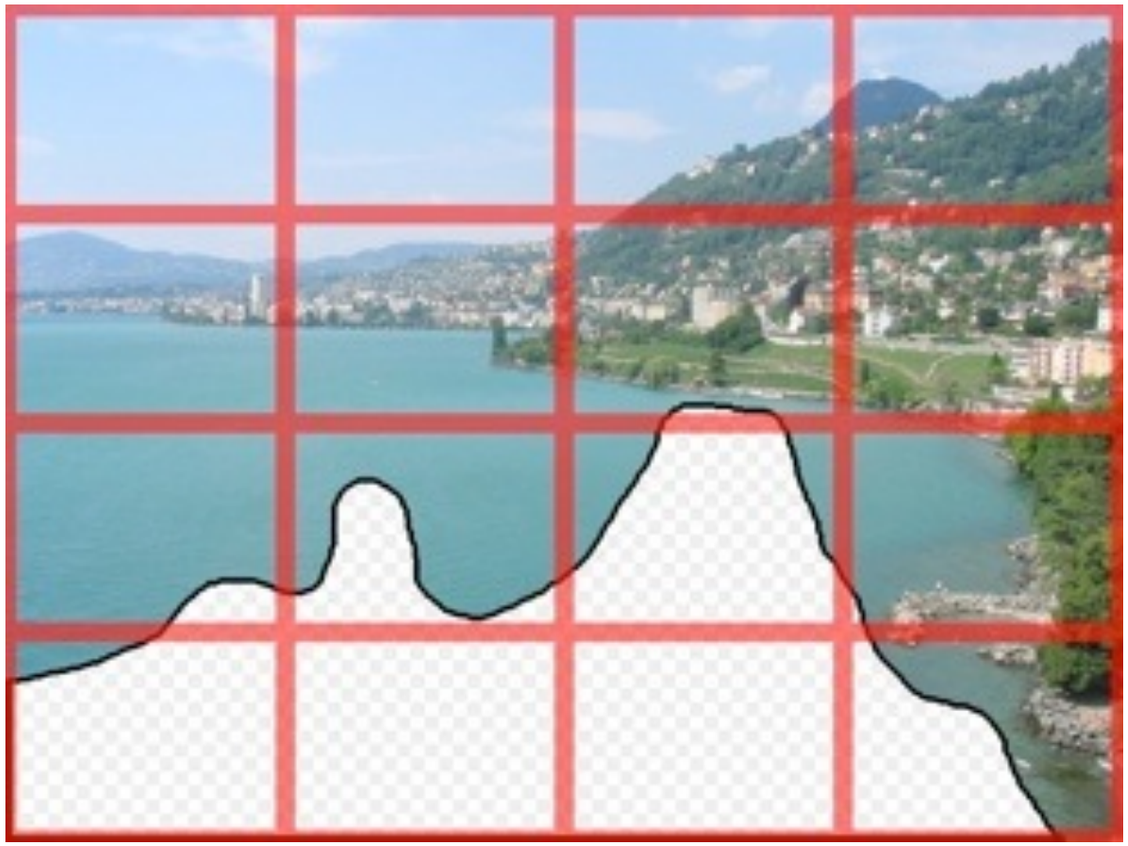
Descripteur de scène

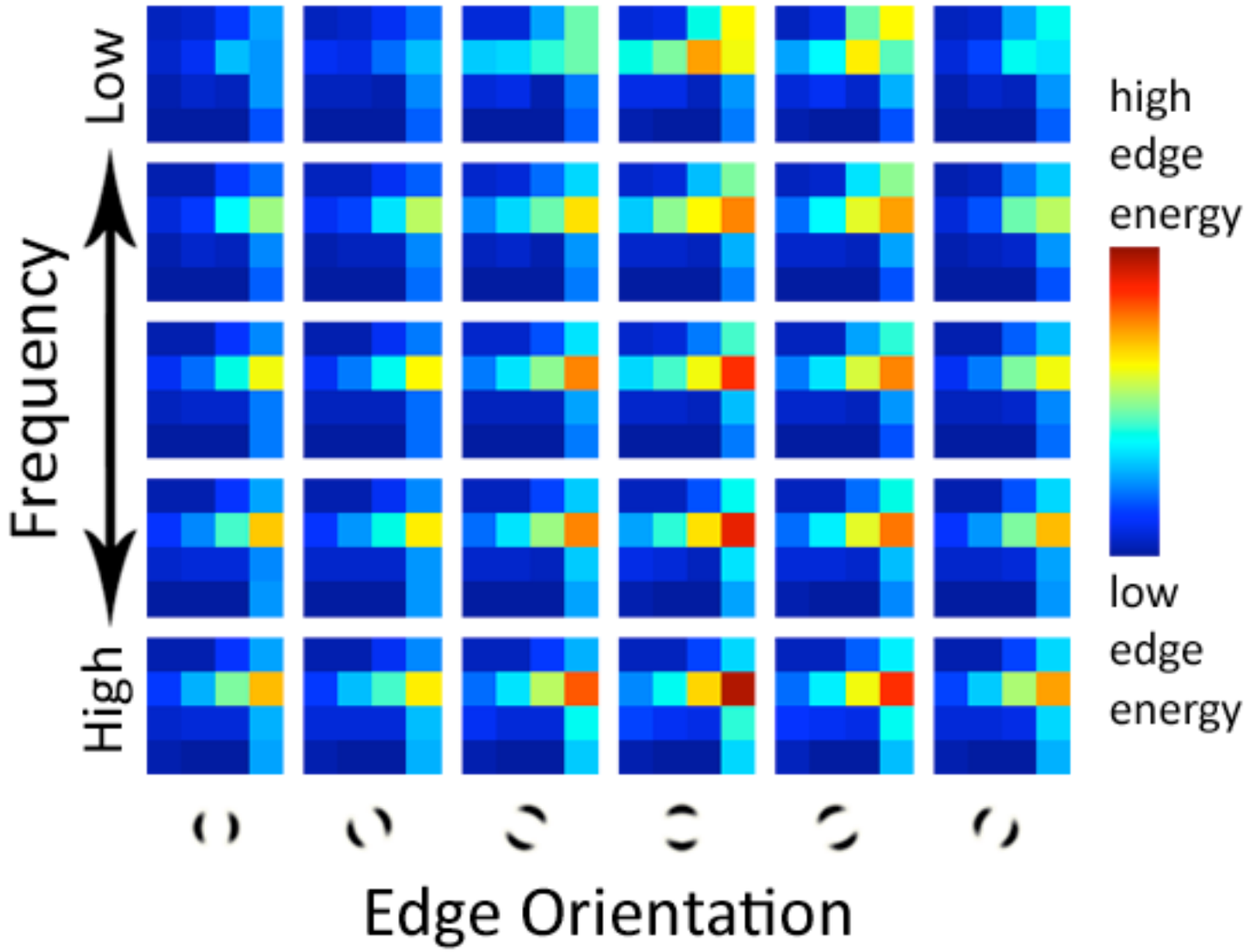




J. Hays

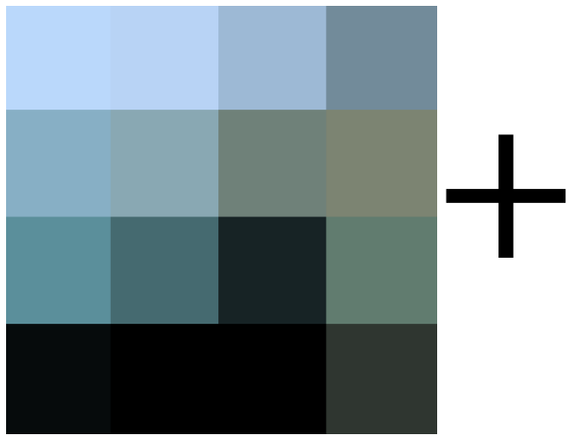
Descripteur de scène



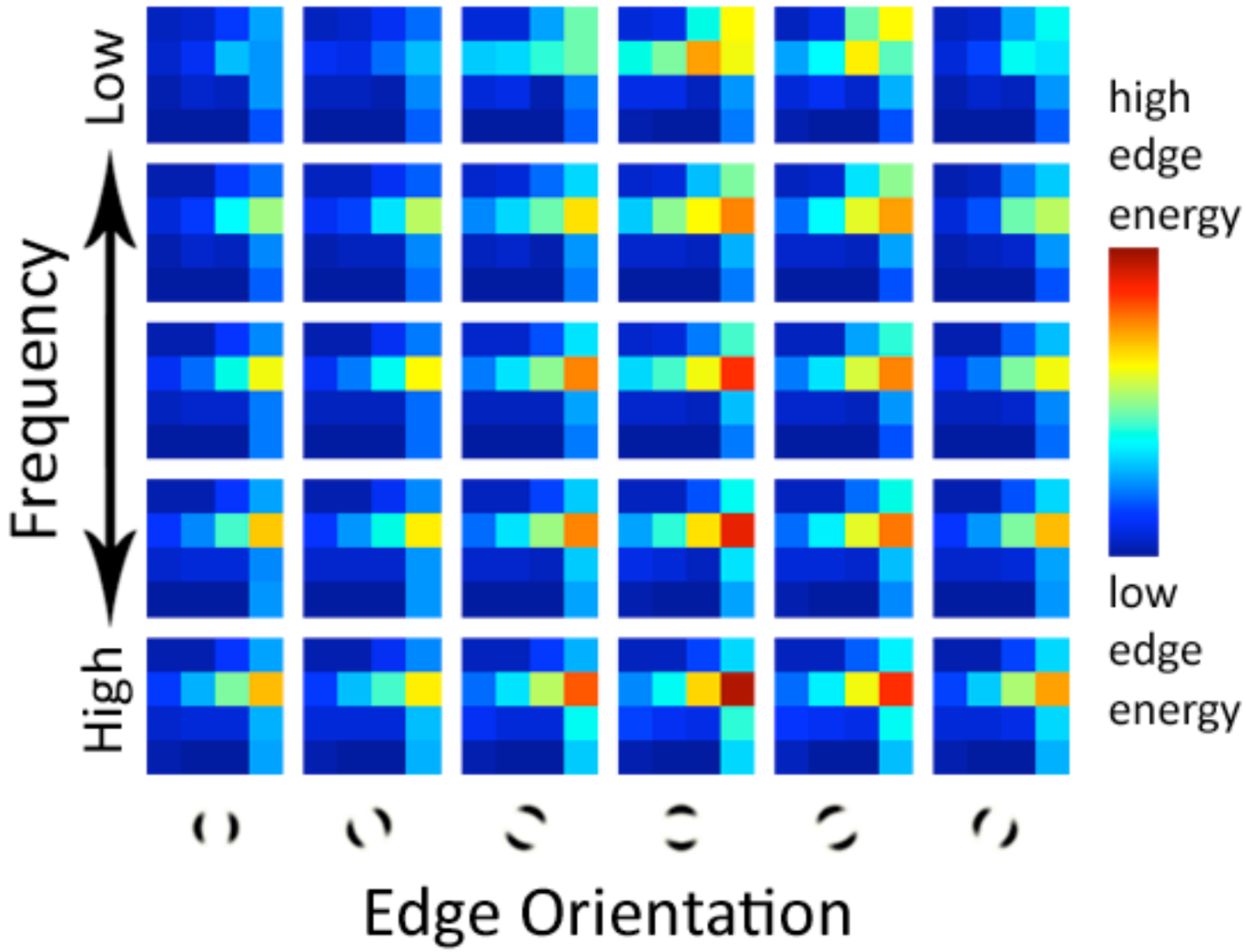


Scene Gist Descriptor
(Oliva and Torralba 2001)

Descripteur de scène



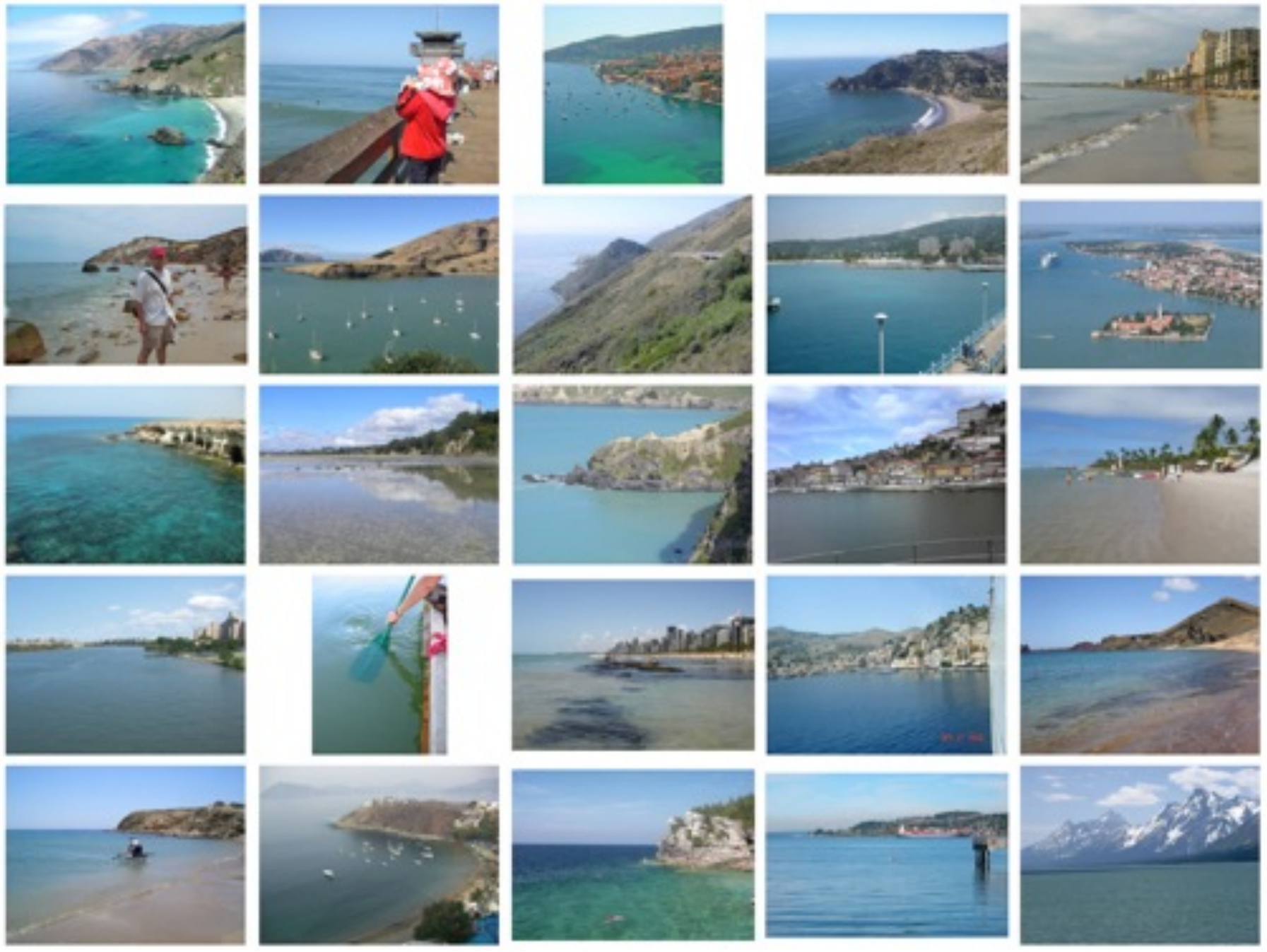
+



Descripteur nommé “gist”
(Oliva and Torralba 2001)

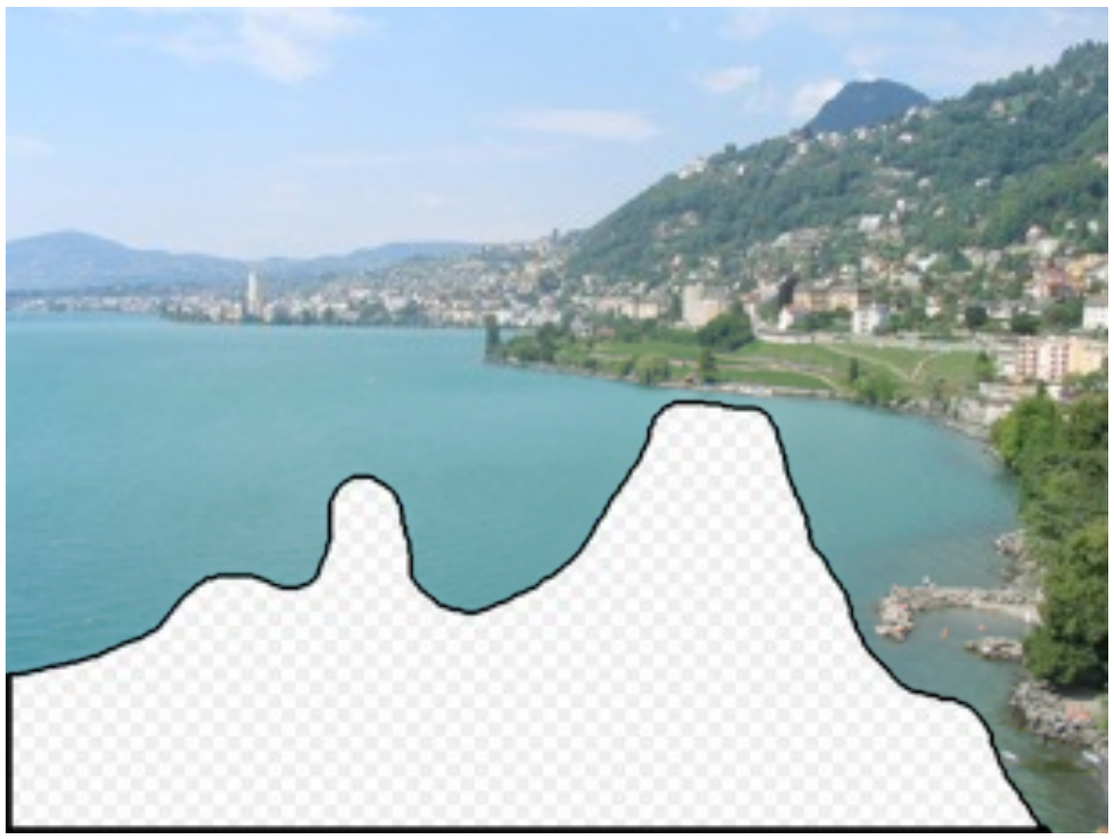
J. Hays

2 millions d’images de Flickr



… 200 total

Appariement local

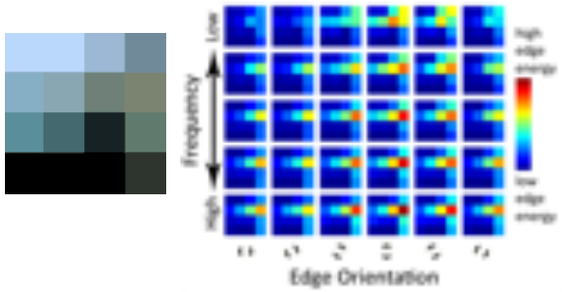


Graph cut + mélange par gradients

J. Hays

Ordonner les résultats

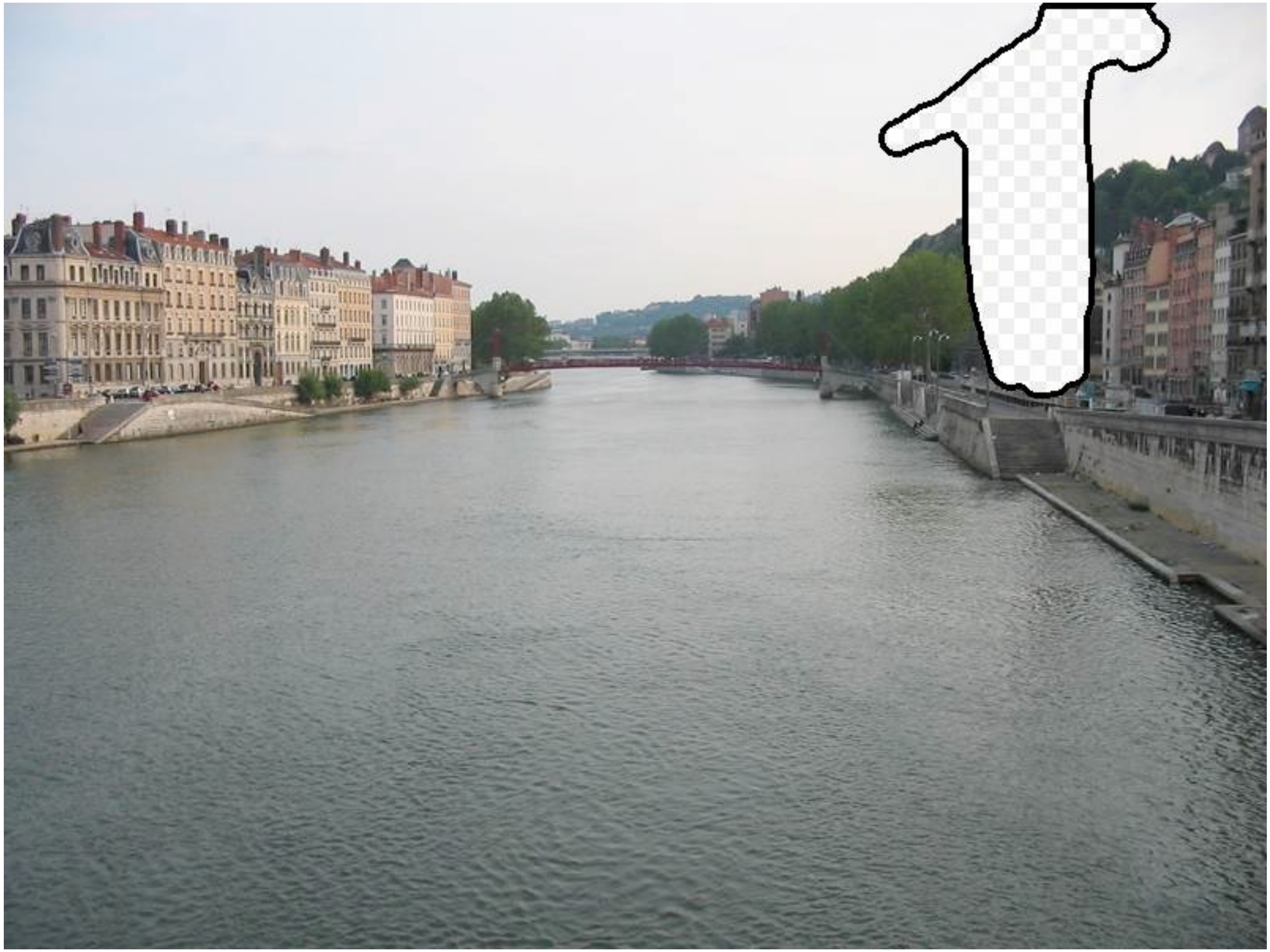
Score final est la somme de:



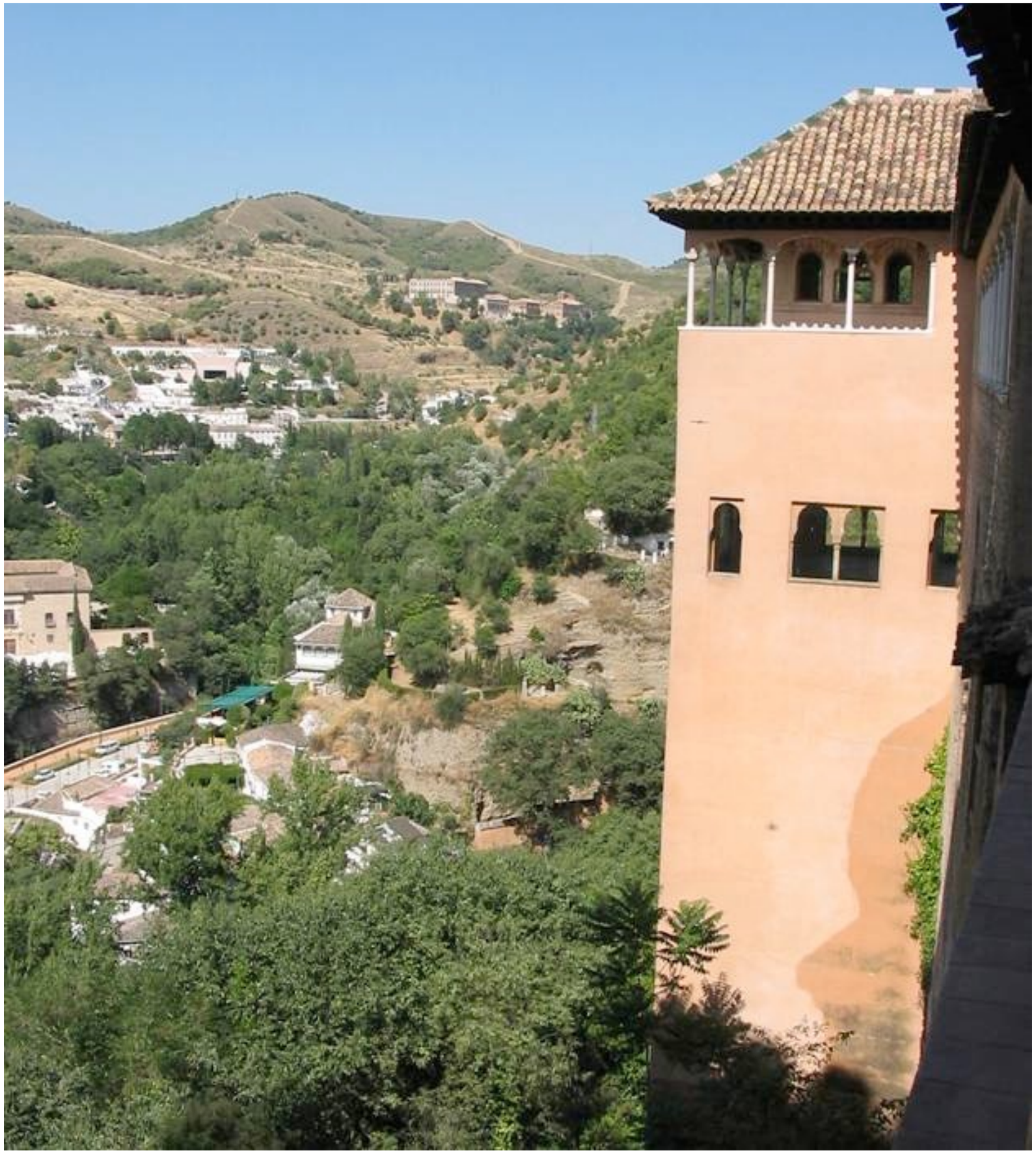
L’appariement de scènes

L’appariement local (couleur + texture)

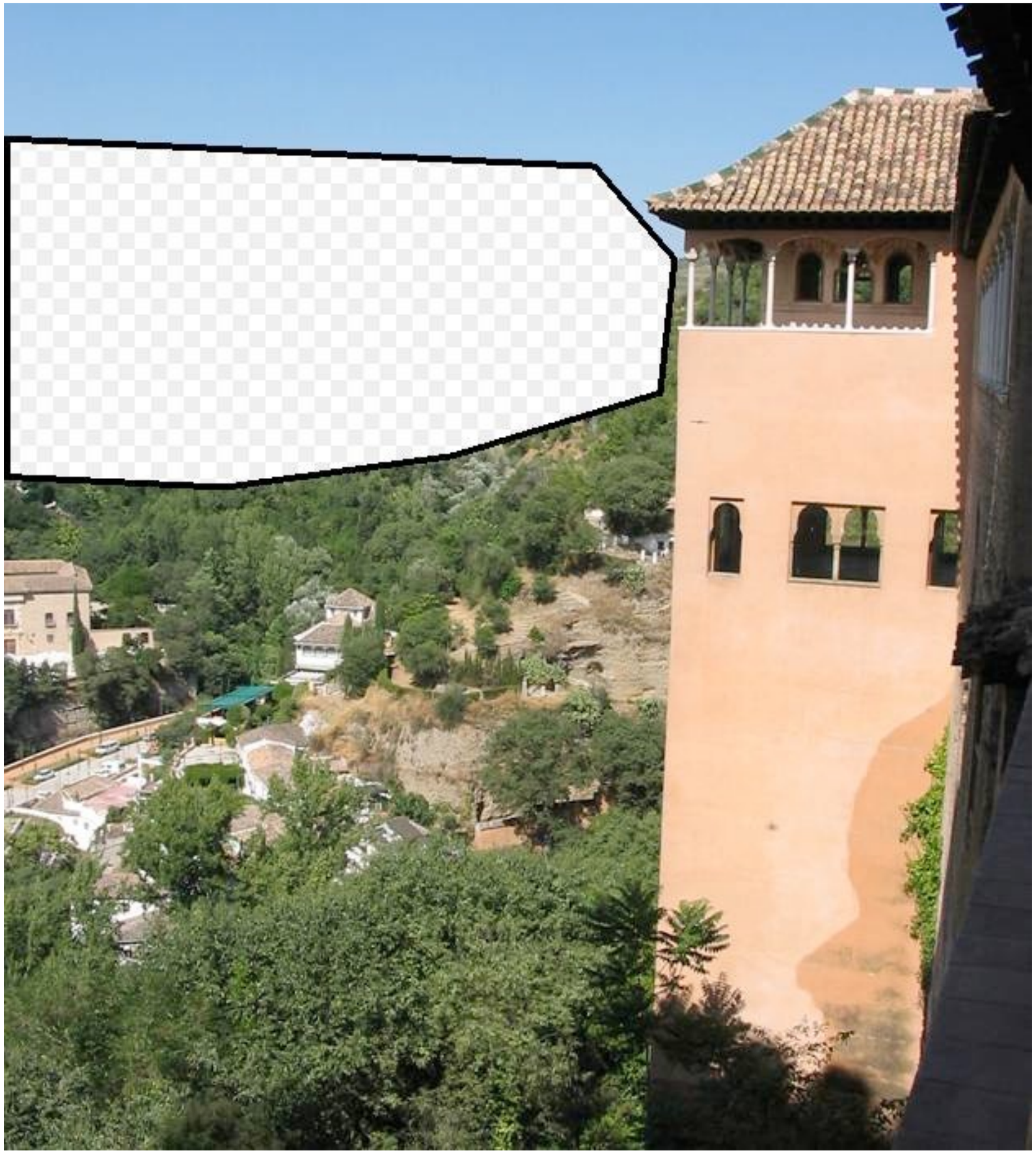
Le coût de la coupure de graphe



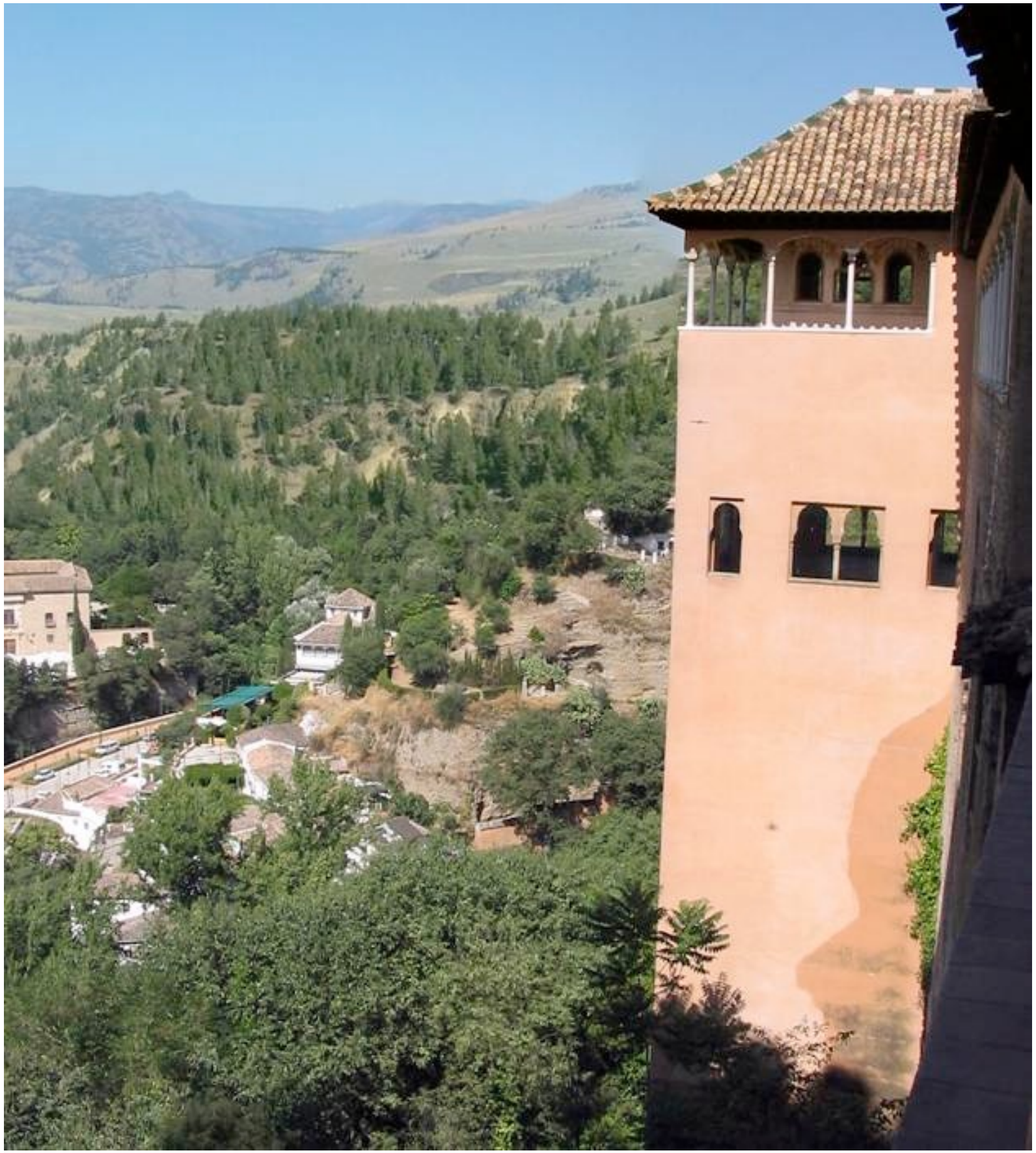
… 200 scenes



J. Hays



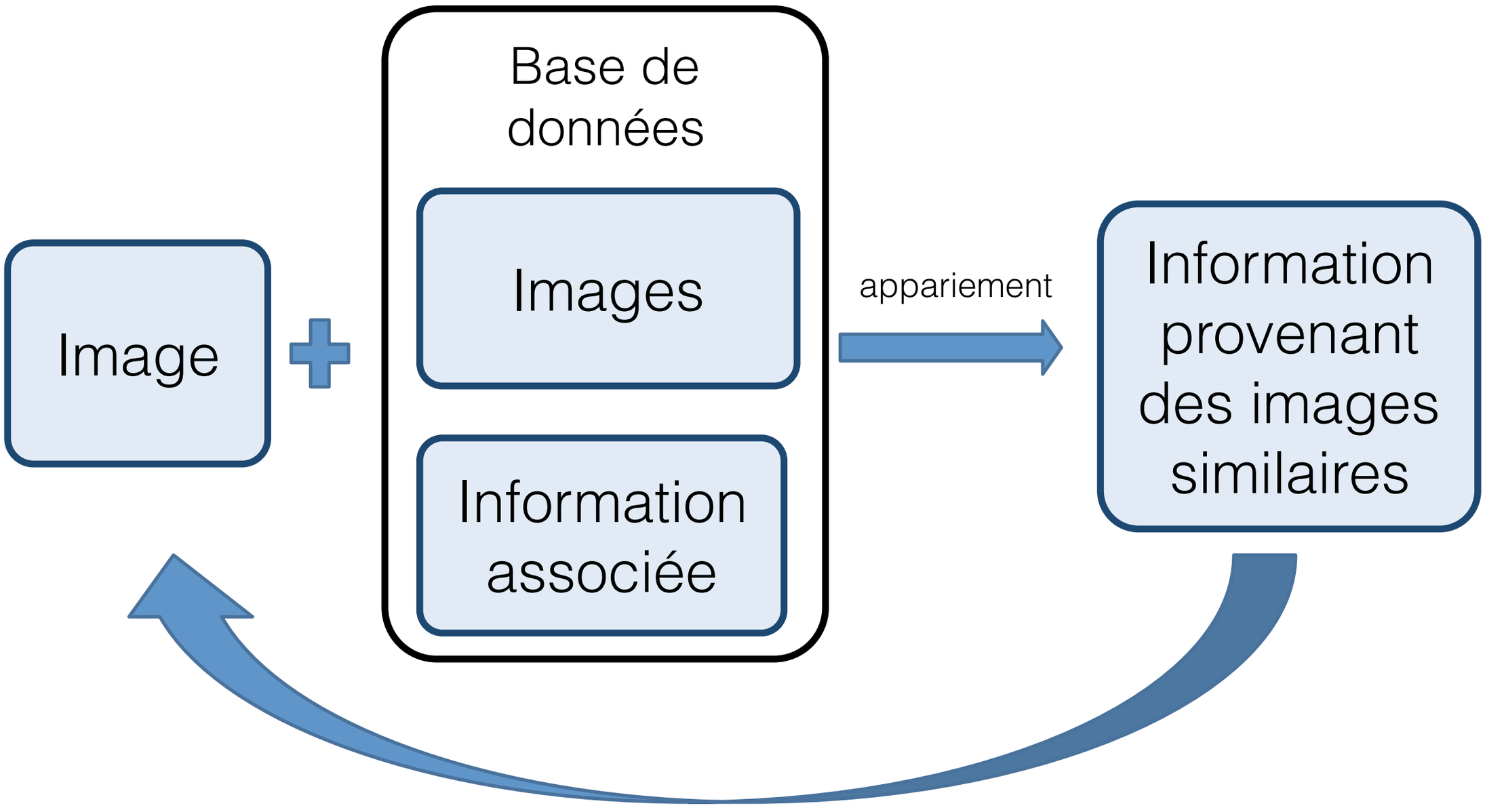
J. Hays



J. Hays

J. Hays

Utiliser beaucoup de données!



Truc: si vous avez assez d'images, la base de données devrait contenir des images suffisamment similaires, faciles à trouver!

D. Hoiem

Combien d’images?



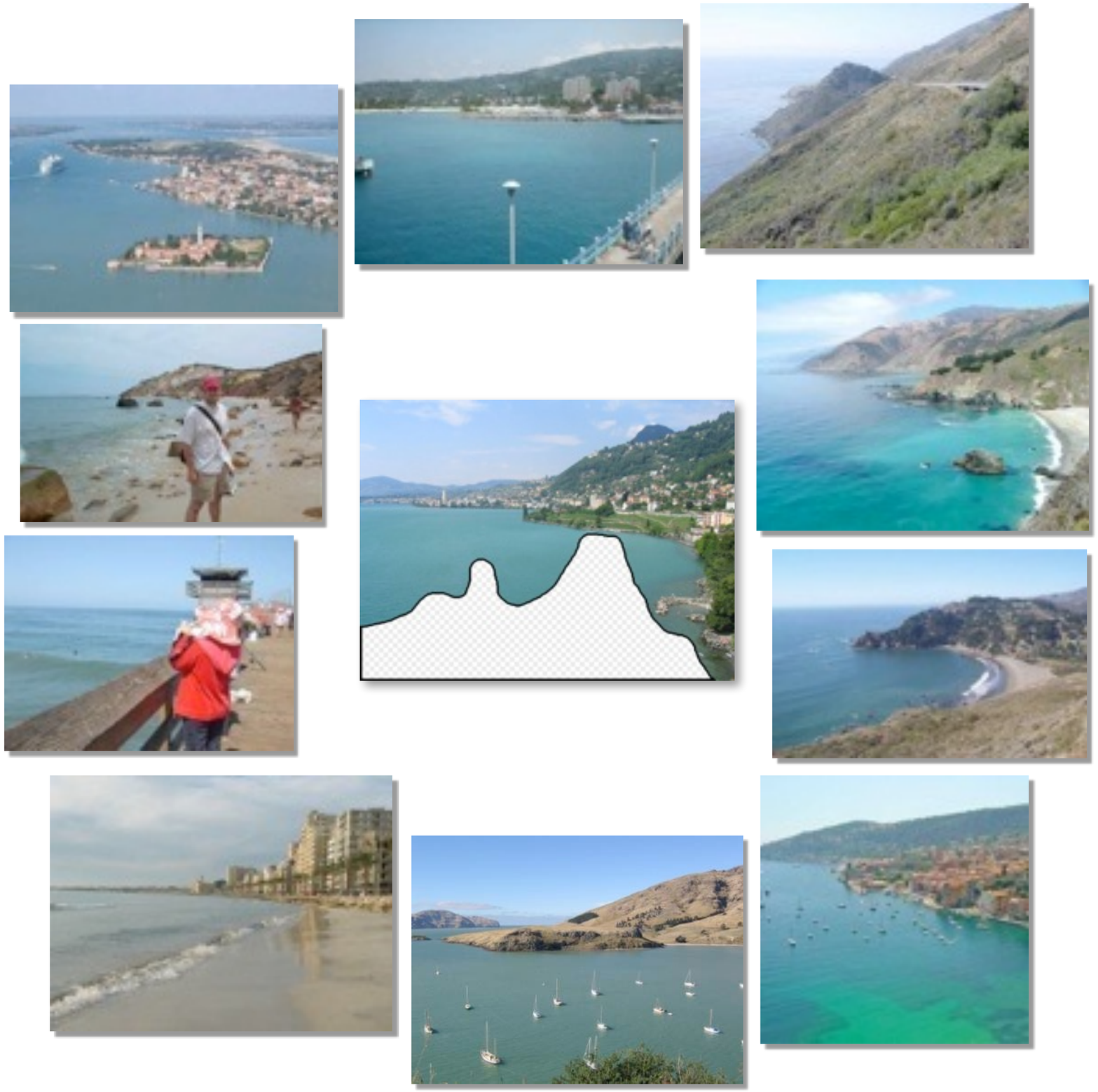








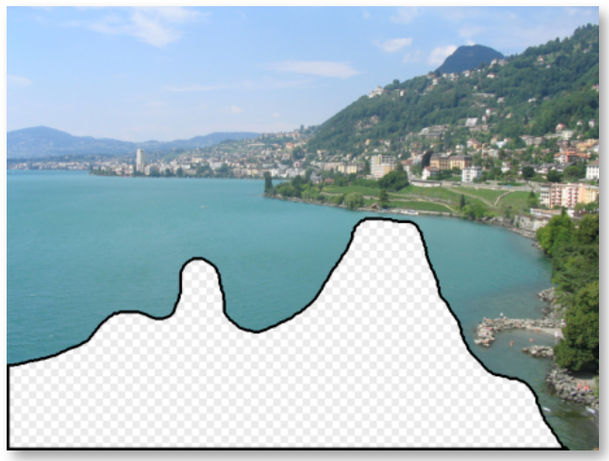
20,000 images

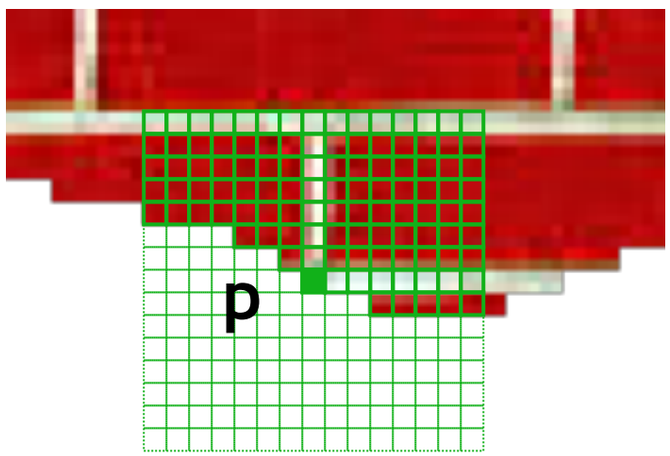


2,000,000 images

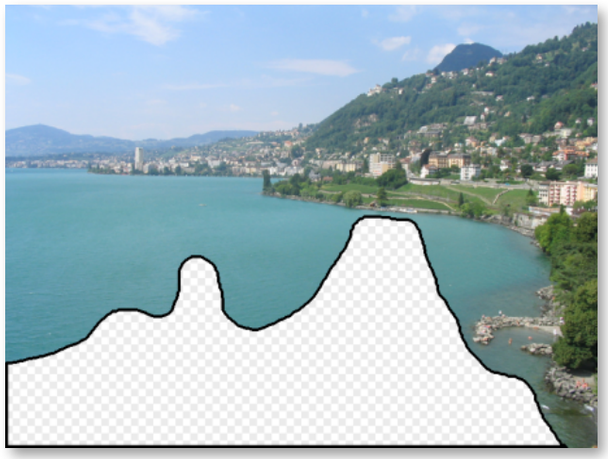
J. Hays

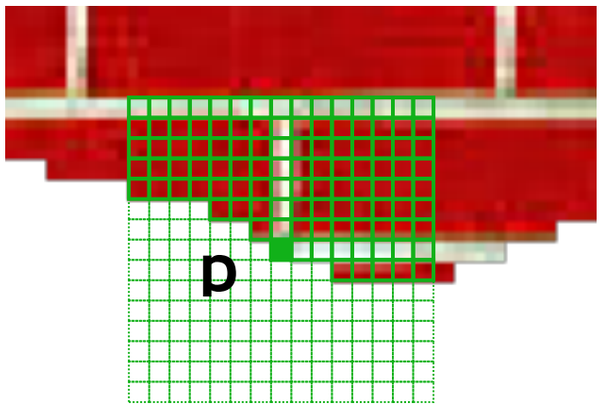
Limites





- Lent
- Fonction d'appariement définie manuellement (peut sembler arbitraire)
- Chaque méthode fonctionne seulement pour leur domaine particulier (e.g. scènes «typiques» à l'extérieur, textures semi-régulières, etc.)



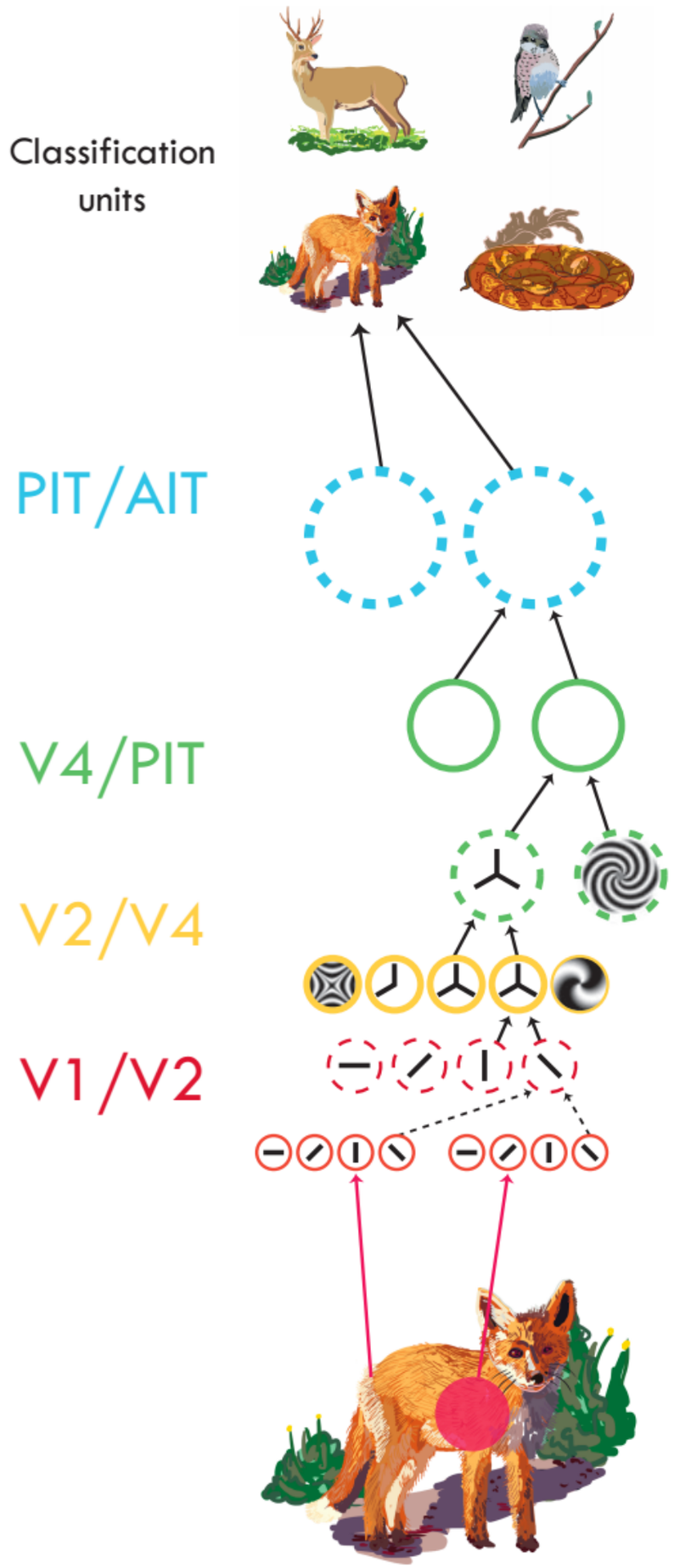


Stocke toute l’information dans une mémoire

«lookup table»

Assume que le monde possède une structure simple

Apprend une représentation qui capture cette structure



Serre, 2014

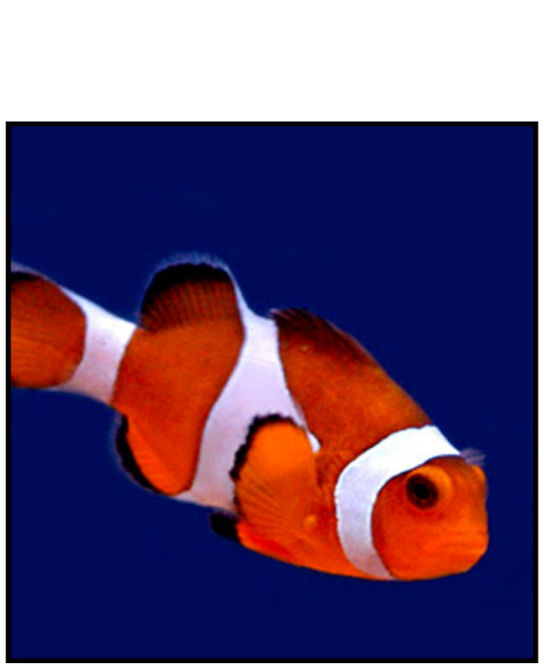
Idée de base

Cerveau/Machine

«poisson clown»

Hiérarchie d’unités *simples*

Reconnaissance d'objets: approche «traditionnelle»



Arêtes

Texture

Couleurs

Segments

Parties

«poisson clown»

Calcul de caractéristiques (features)

Classificateur

Reconnaissance d'objets: approche «traditionnelle»

Appris automatiquement

Arêtes

Texture

Couleurs

Segments

Parties

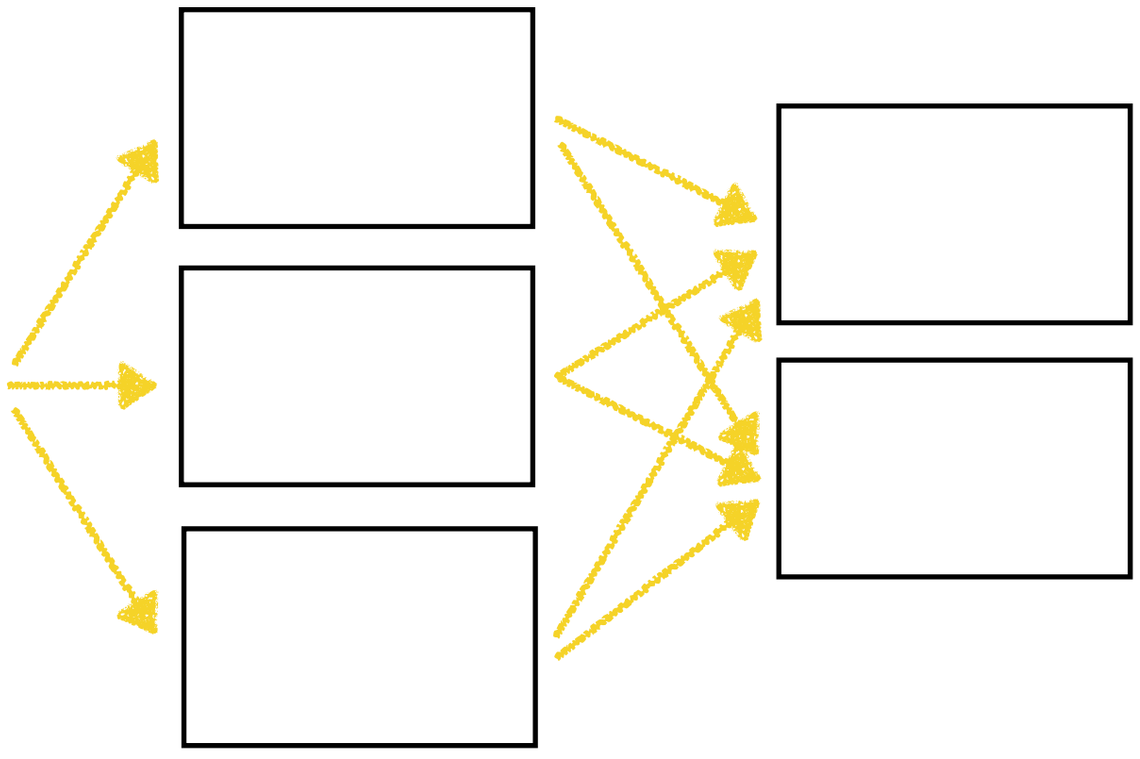
«poisson clown»

Calcul de caractéristiques (features)

Classificateur

Réseau de neurones

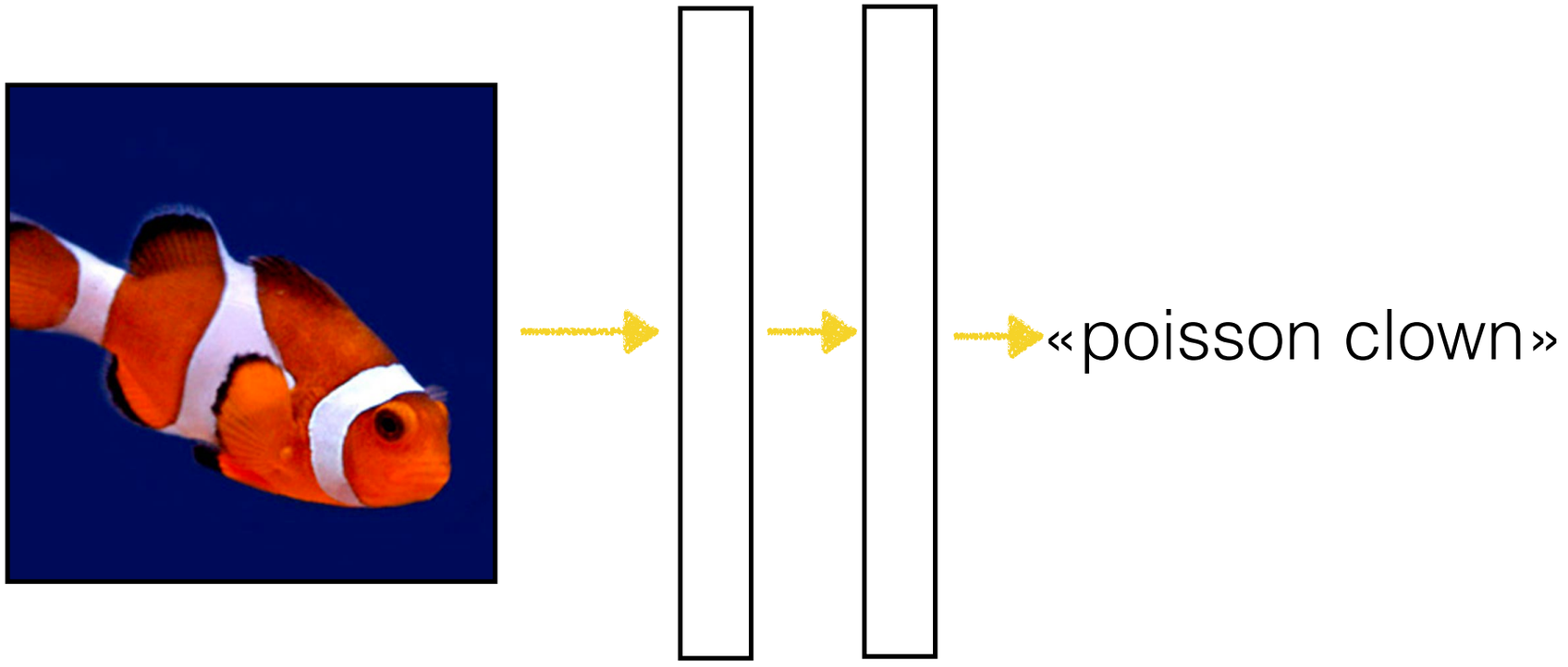
Appris automatiquement



«poisson clown»

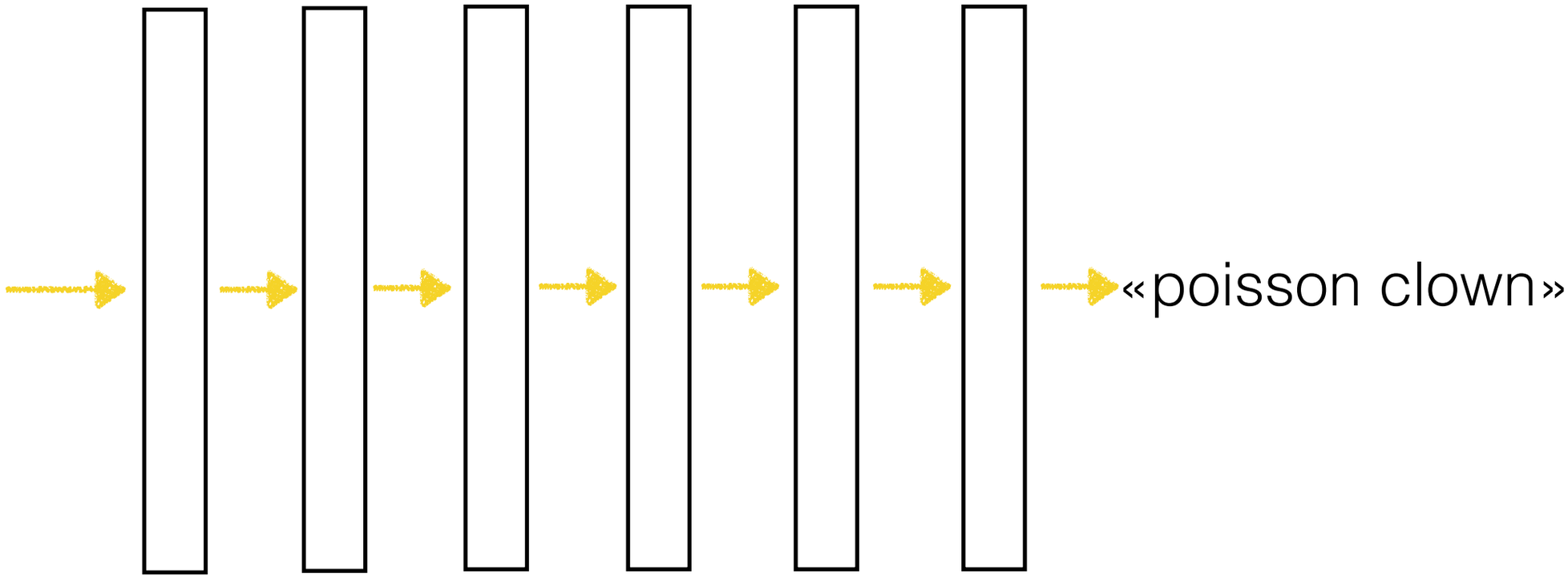
Réseau de neurones

Appris automatiquement



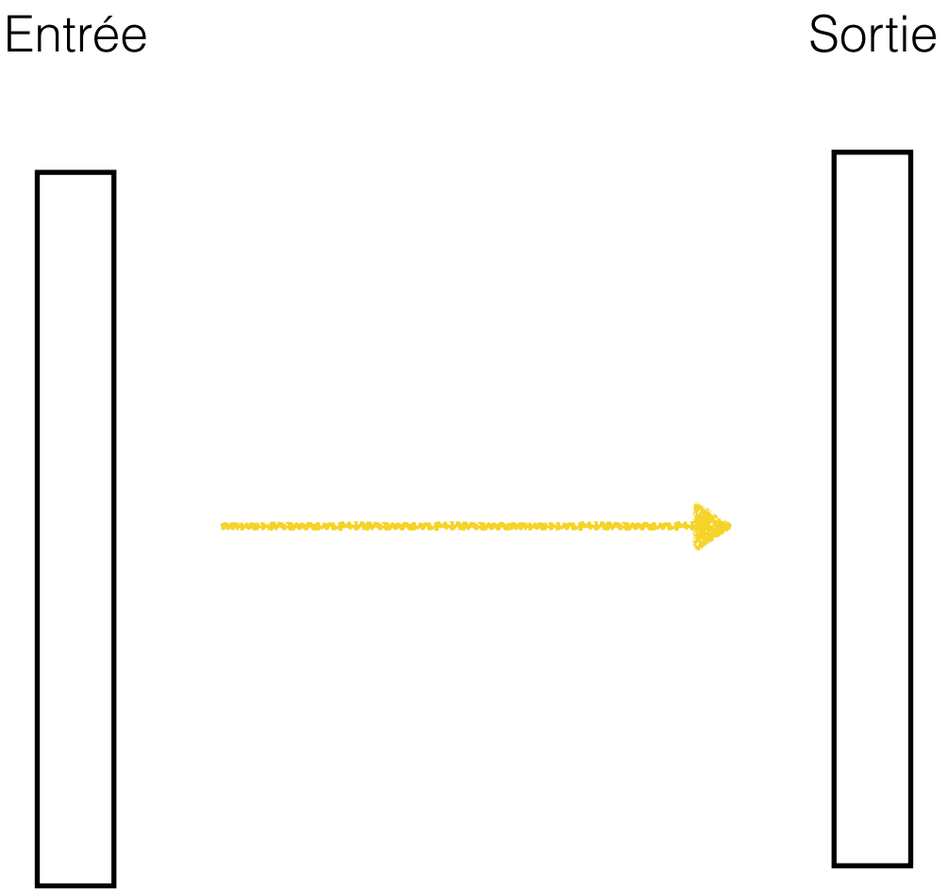
Réseau de neurones *profond*

Appris automatiquement

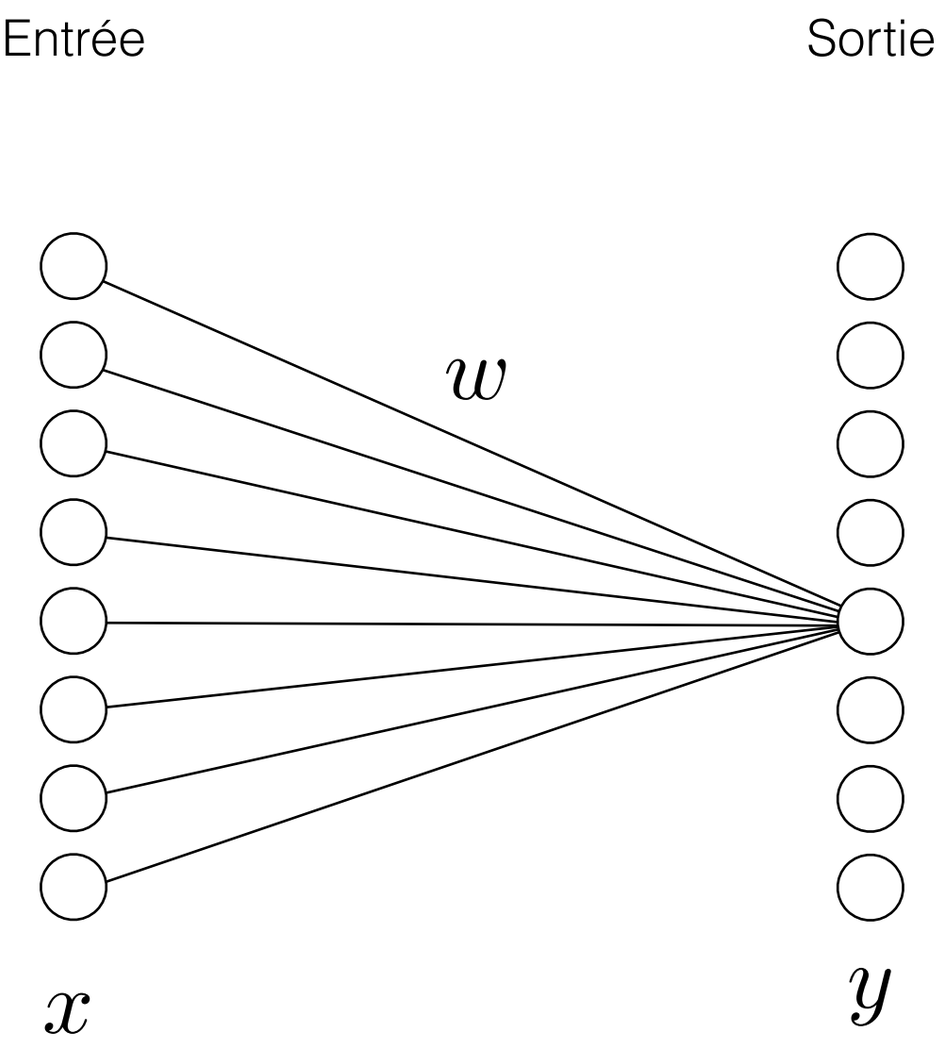


Calculs dans un réseau de neurones

Que font ces flèches?
Quels calculs sont effectués?

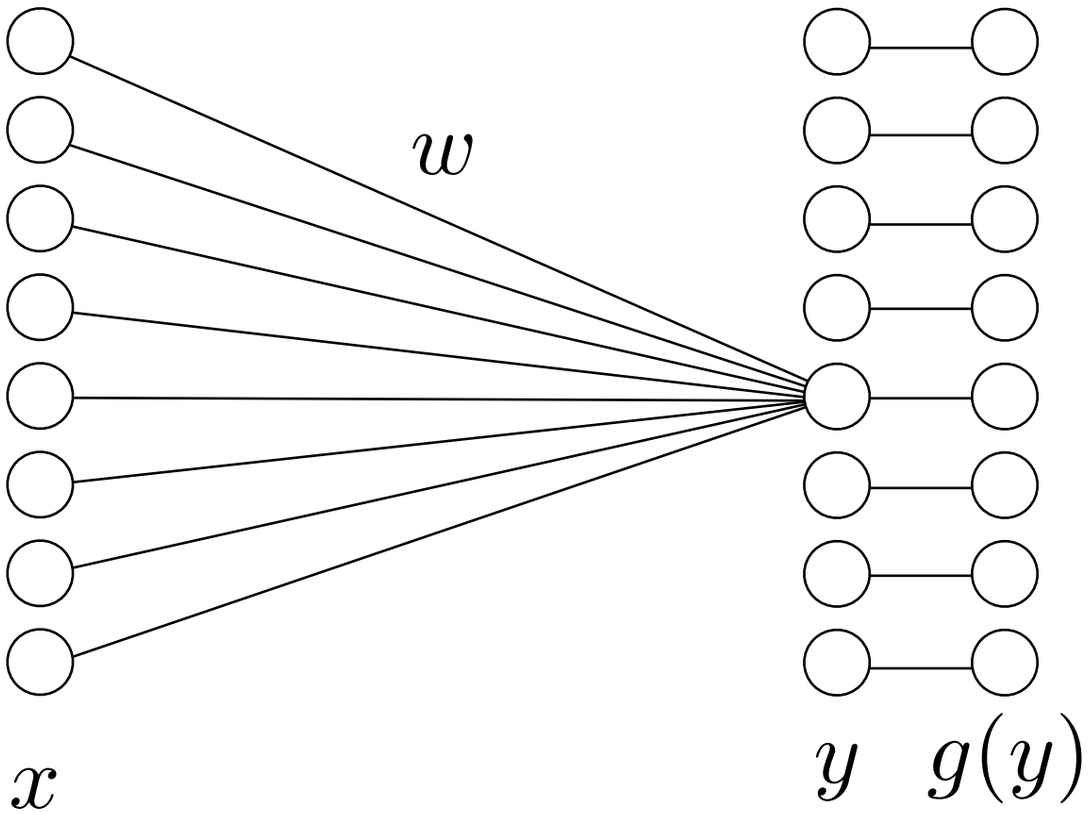


Calculs dans un réseau de neurones

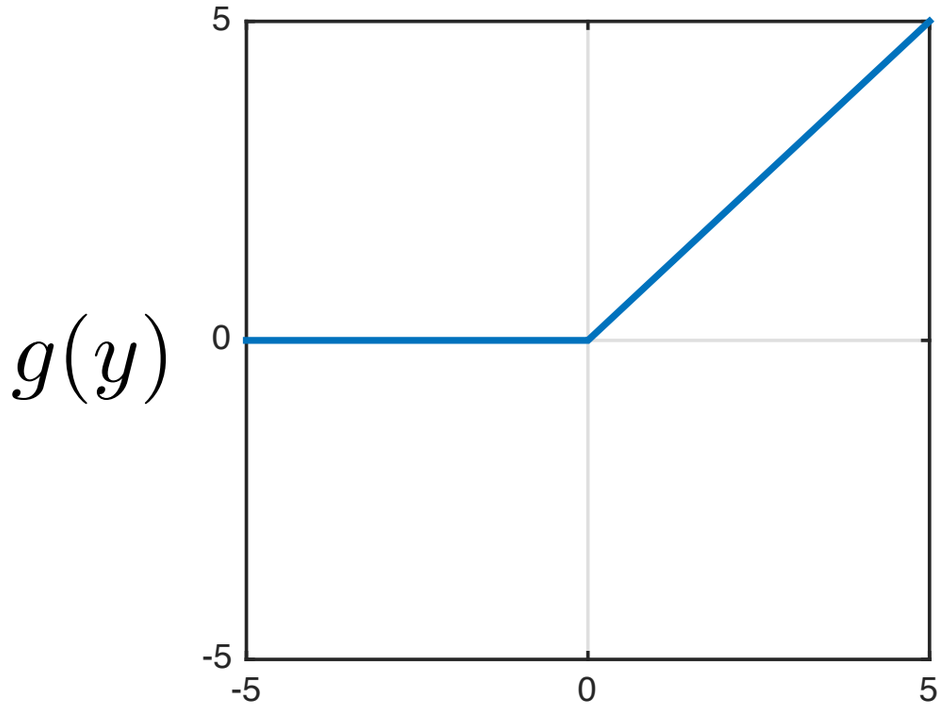


$$y_j = \sum_i w_{ij} x_i$$

Calculs dans un réseau de neurones

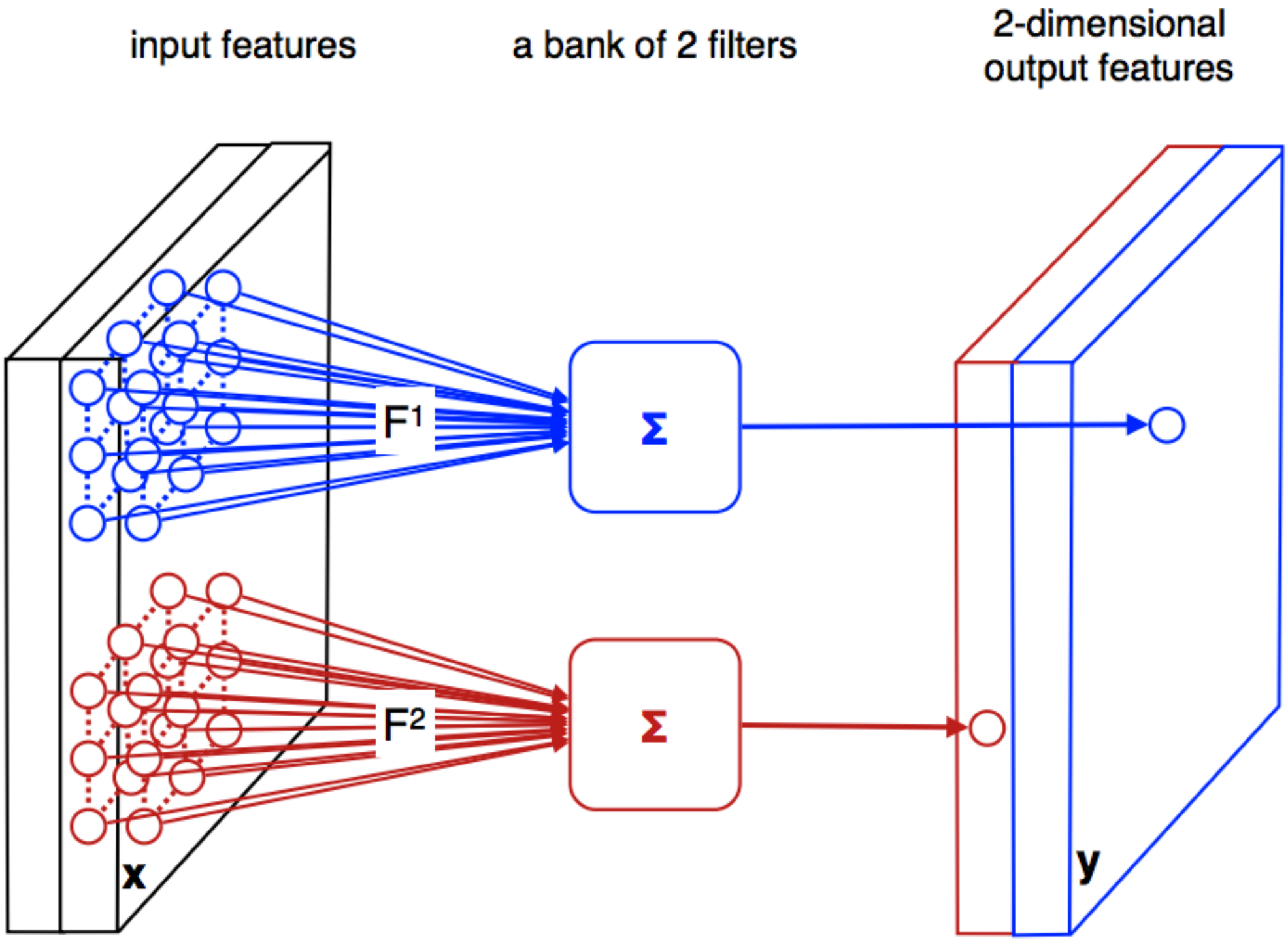


«Rectified linear unit» (ReLU)



$$g(y) = \max(0, y)$$

Réseaux de neurones *à convolution* (CNN)

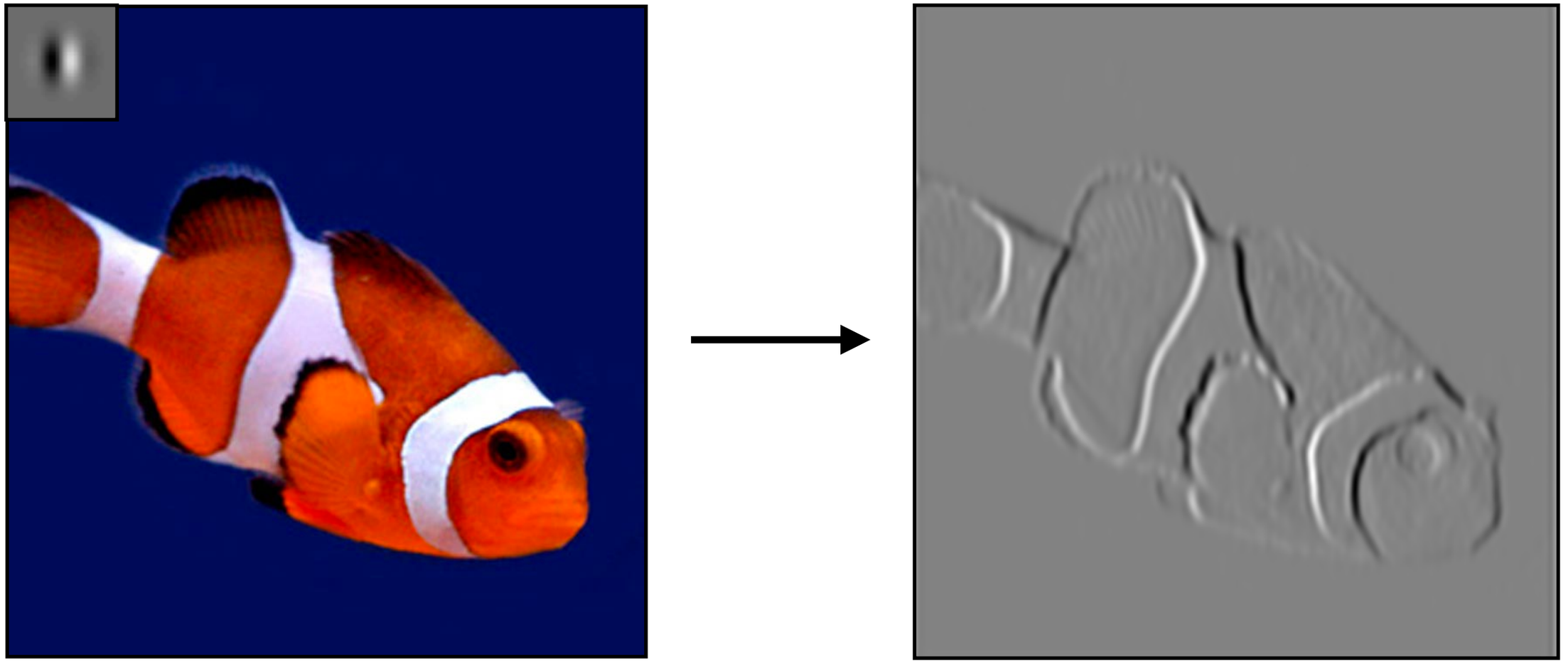


A. Vedaldi

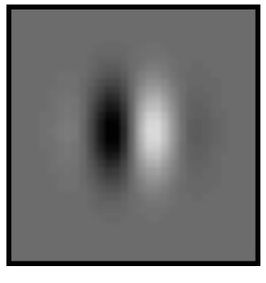
Réseaux de neurones à convolution (CNN)

Convolution

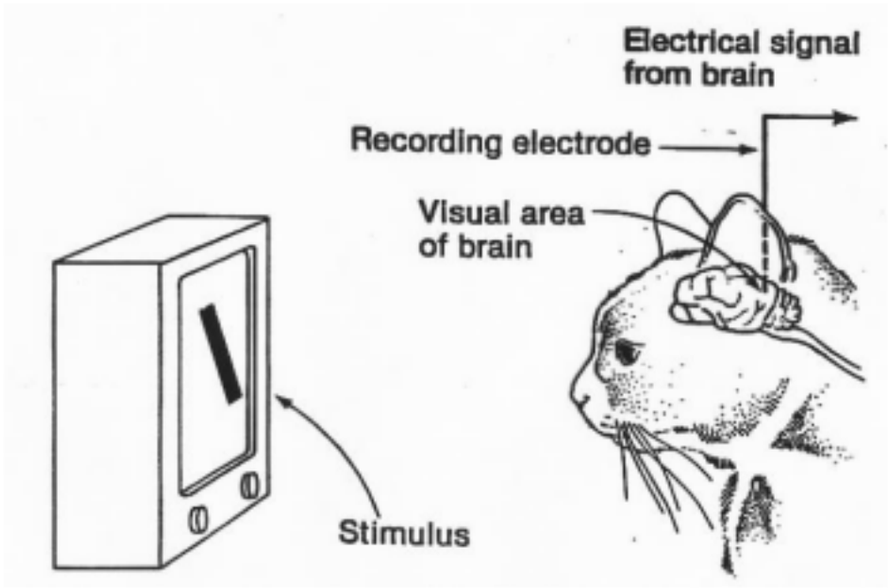
Avantage: les paramètres sont *partagés*
(toutes les fenêtres de l'image sont traitées de la même façon)

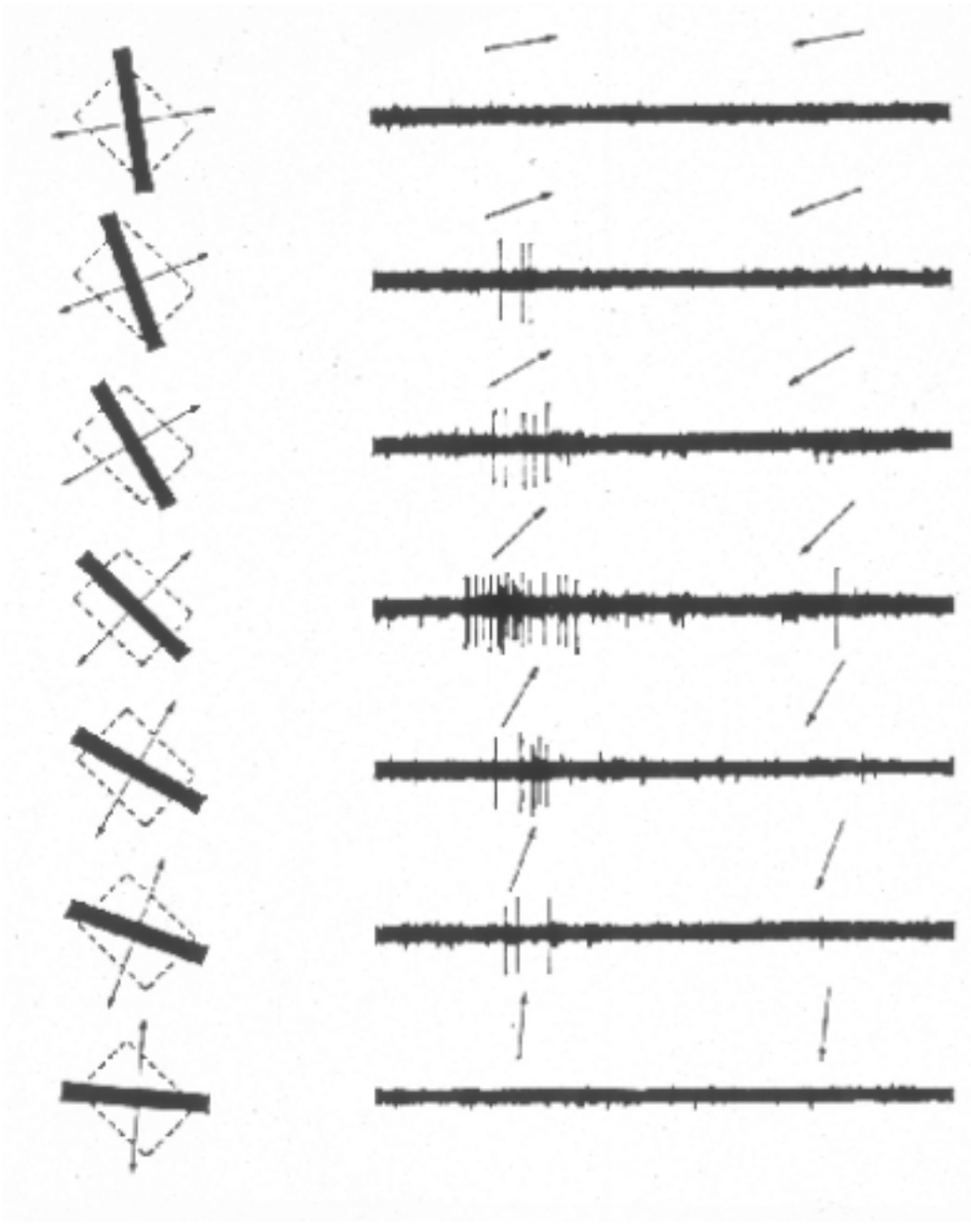


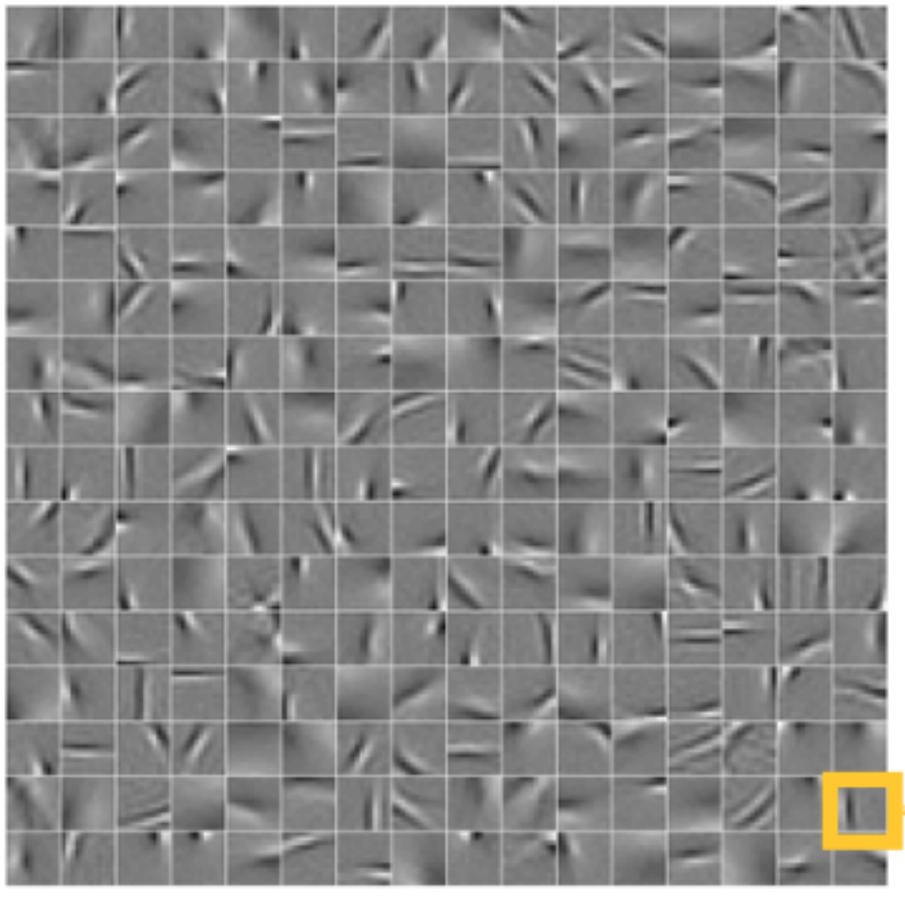
filtre

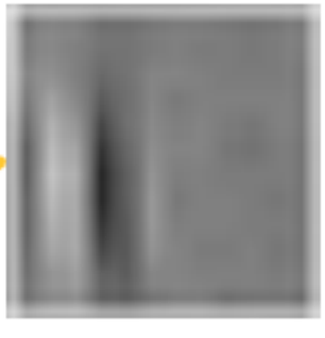


[Hubel and Wiesel 59]



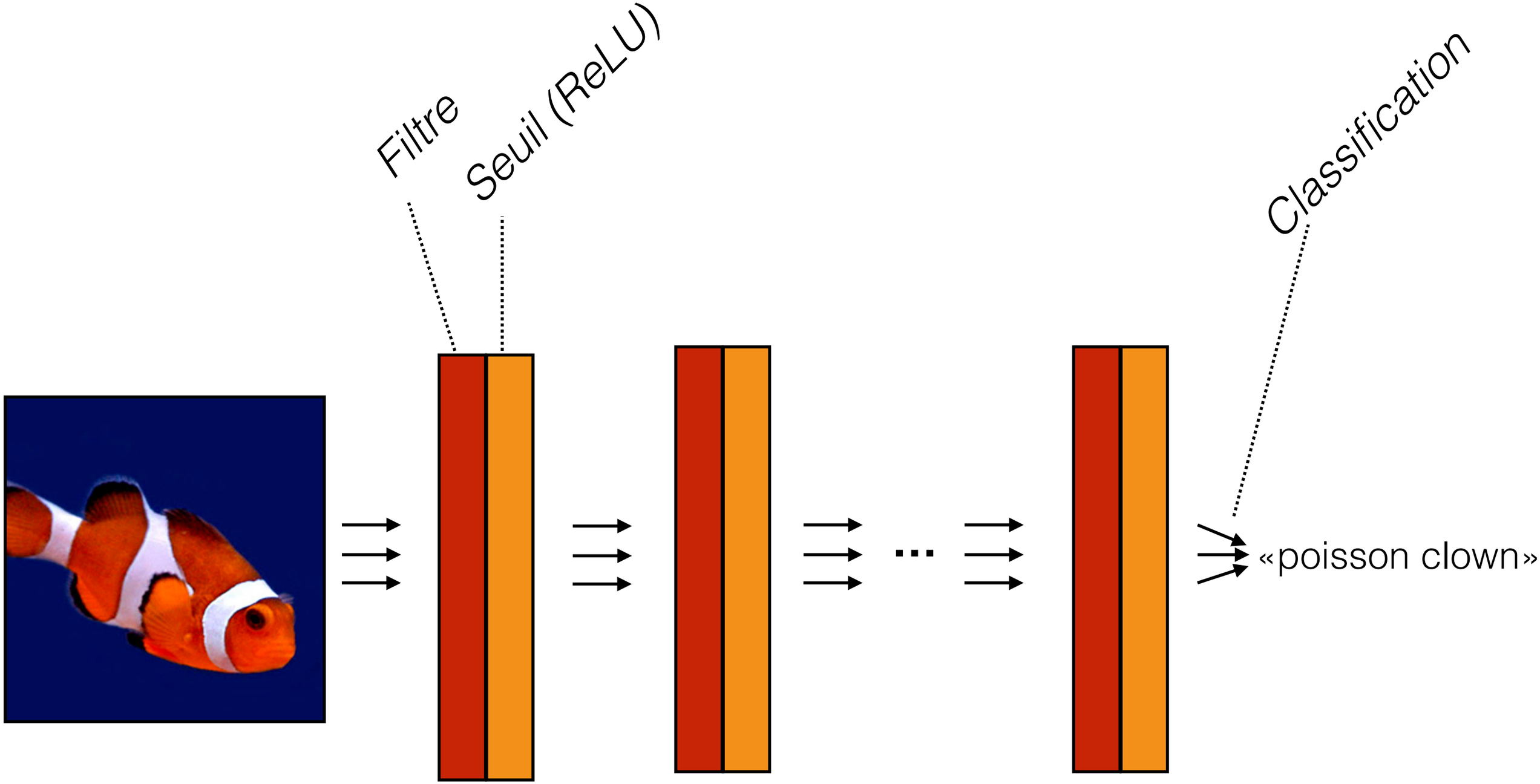






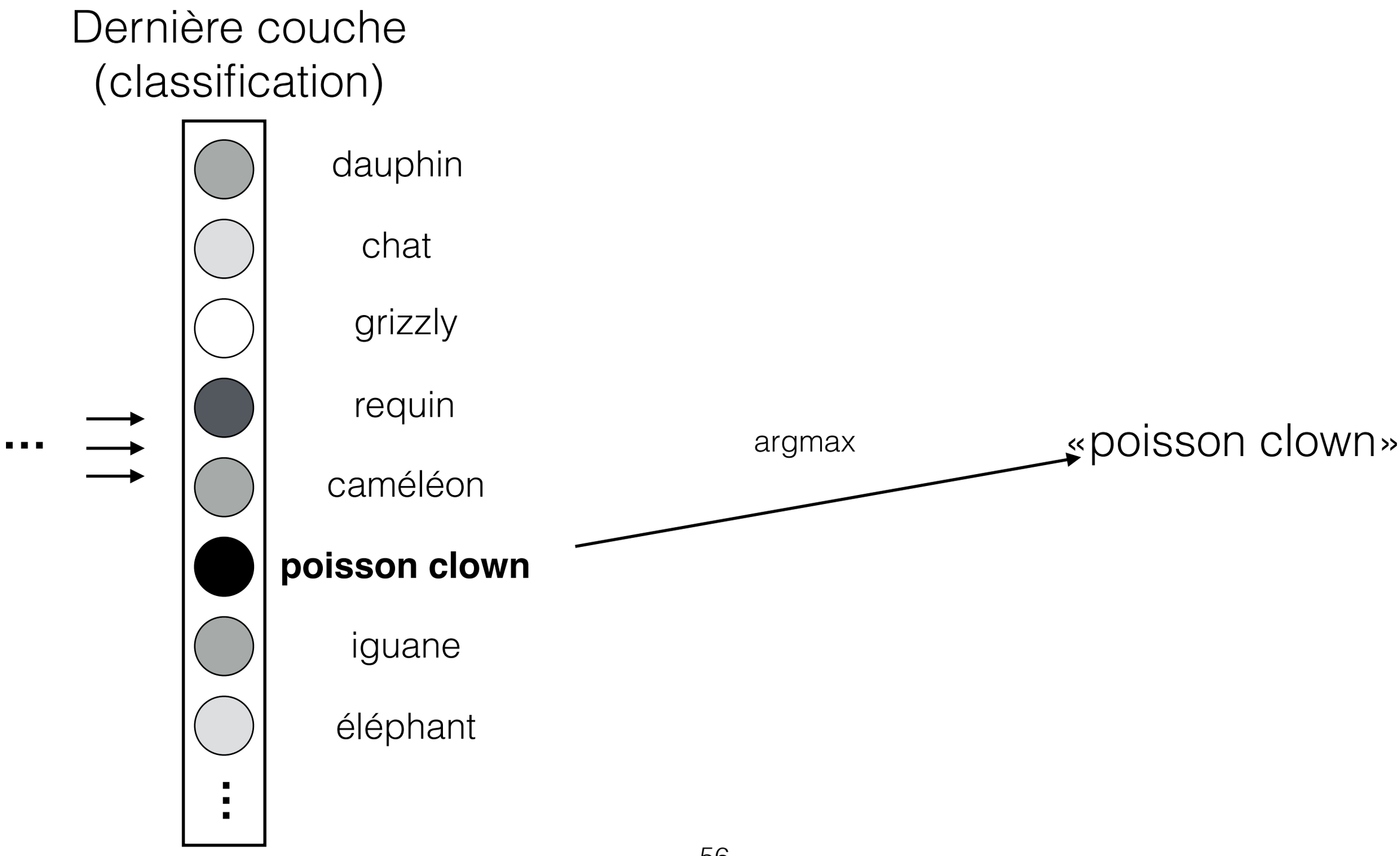
oriented filter

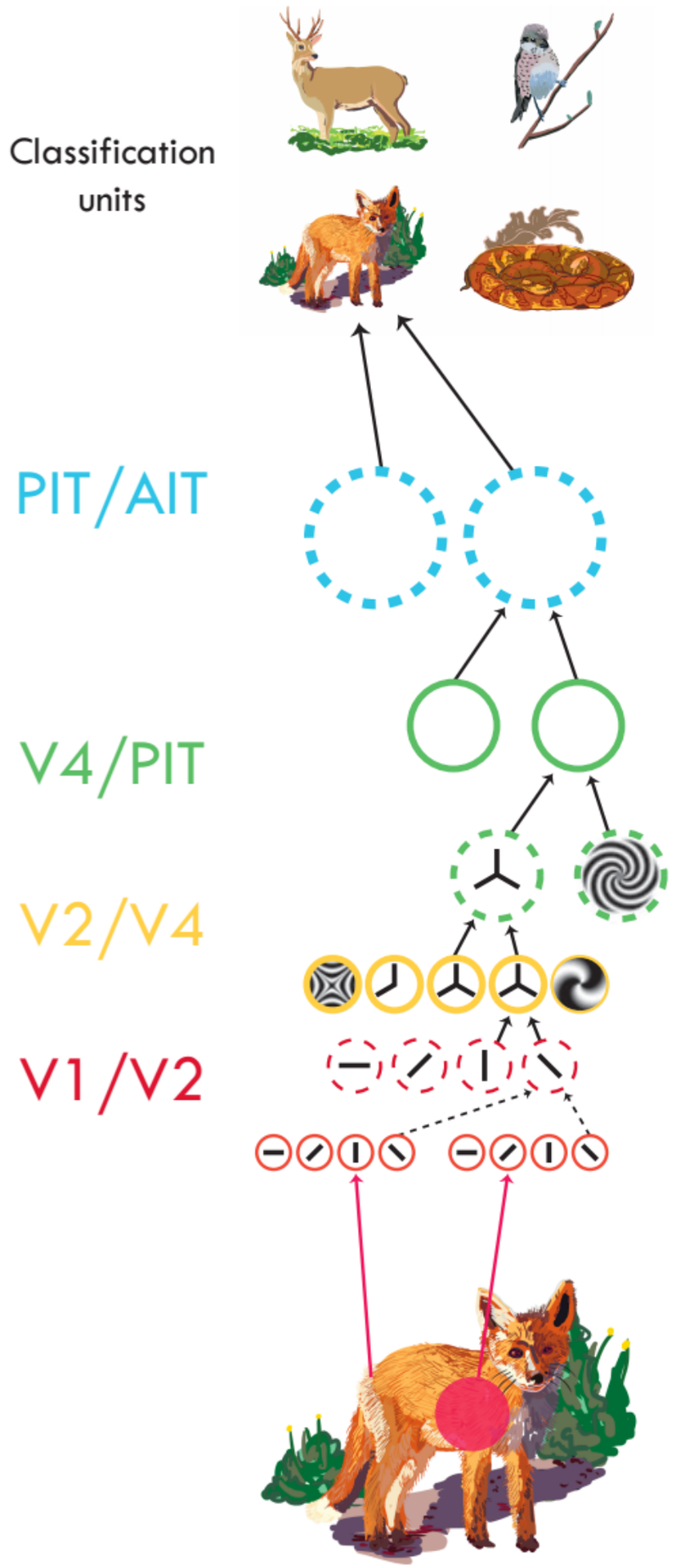
Calculs dans un réseau de neurones



$$f(\mathbf{x}) = f_L(\ldots f_2(f_1(\mathbf{x})))$$

Calculs dans un réseau de neurones

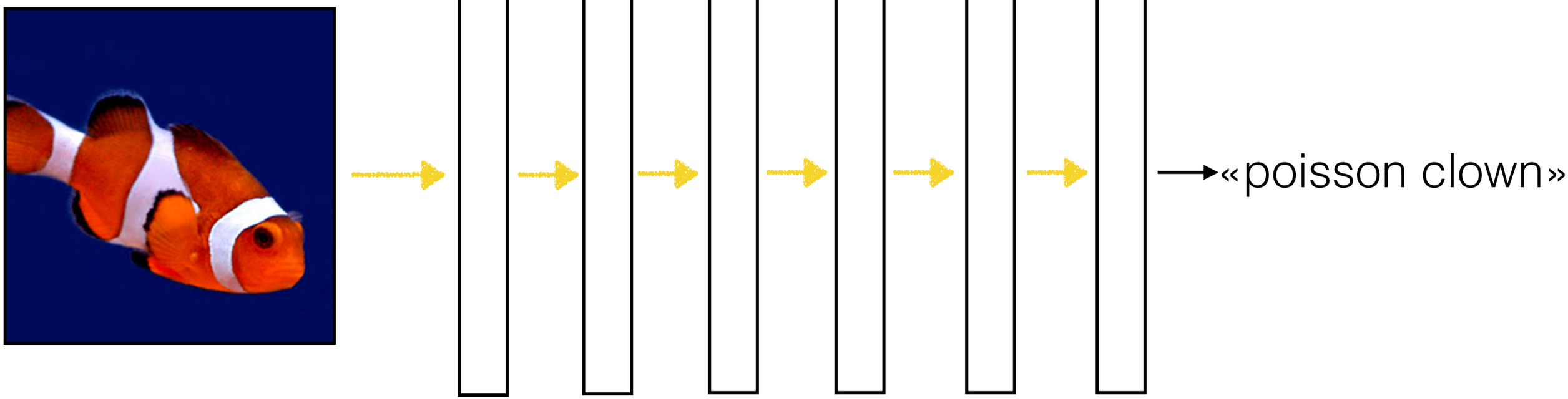




Serre, 2014

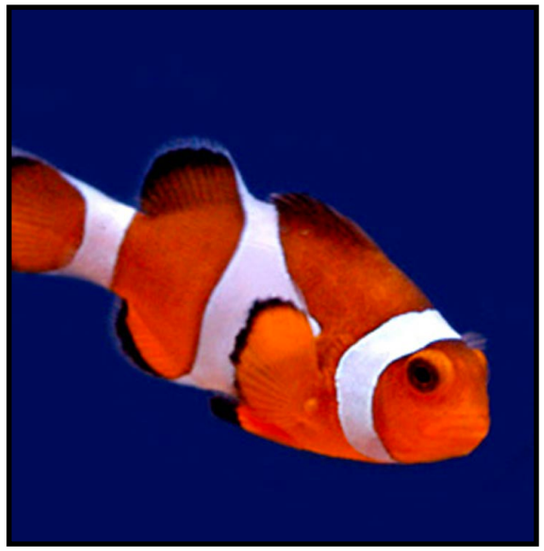
Apprentissage par réseaux profonds

Appris automatiquement



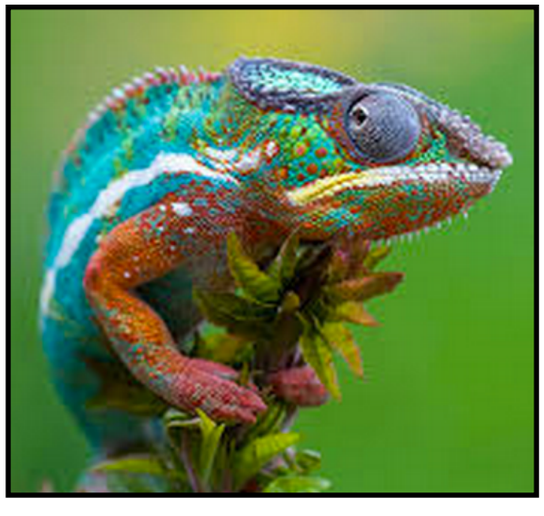
Idée: on modifie les poids (dans un CNN, les filtres!)
jusqu'à temps que la sortie corresponde à ce qu'on veut

Apprentissage par réseaux profonds



→«poisson clown»

→ «grizzly»



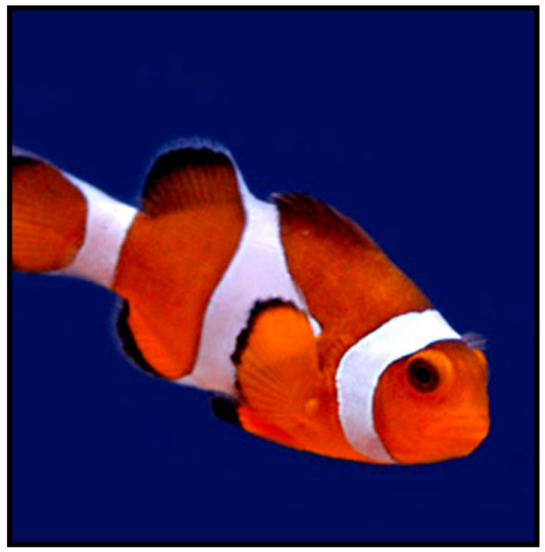
→ «caméléon»

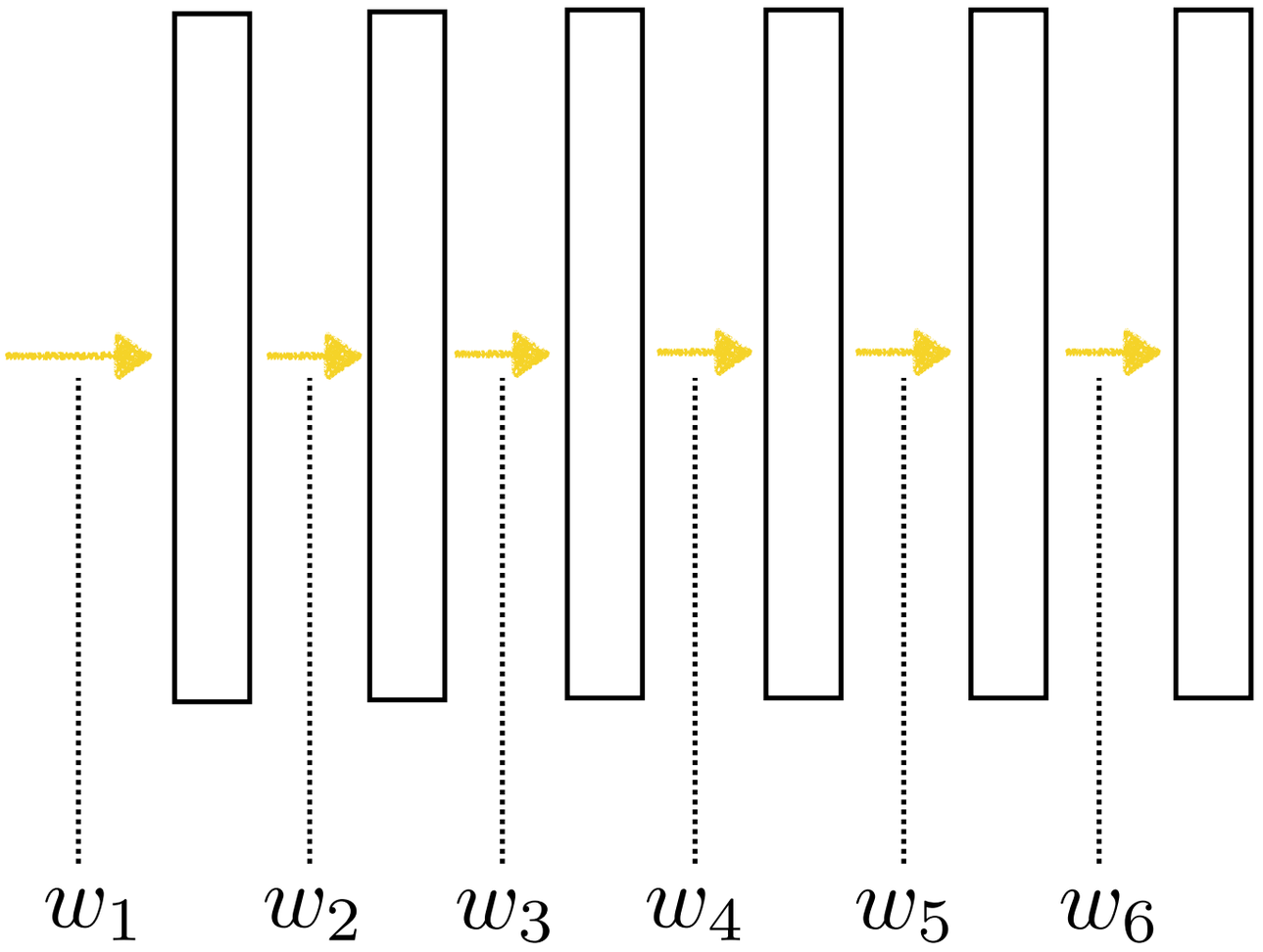
Nous voudrions entraîner le réseau à associer chaque image à la bonne étiquette

Apprentissage par réseaux profonds

Appris automatiquement

«poisson clown»

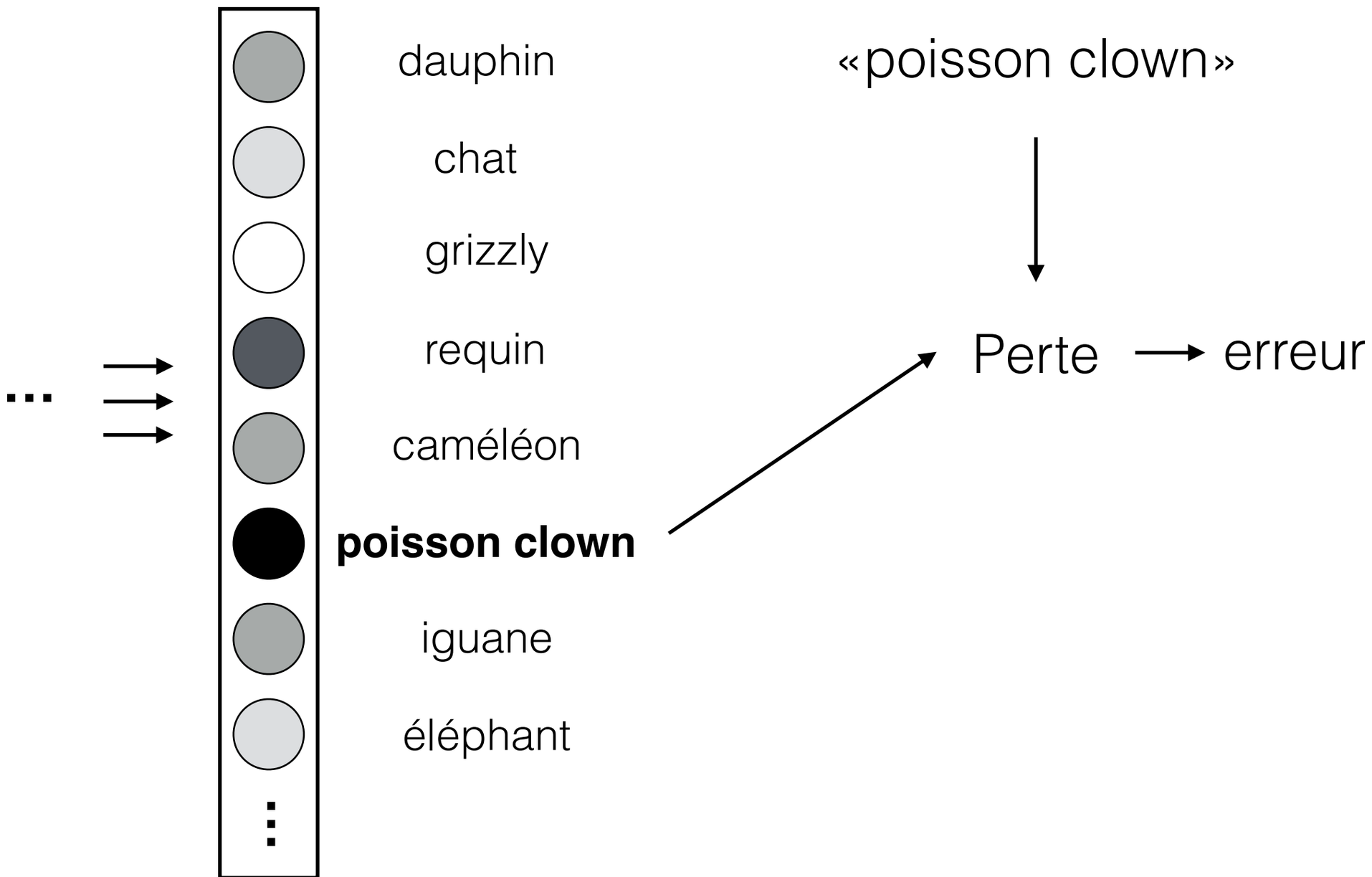




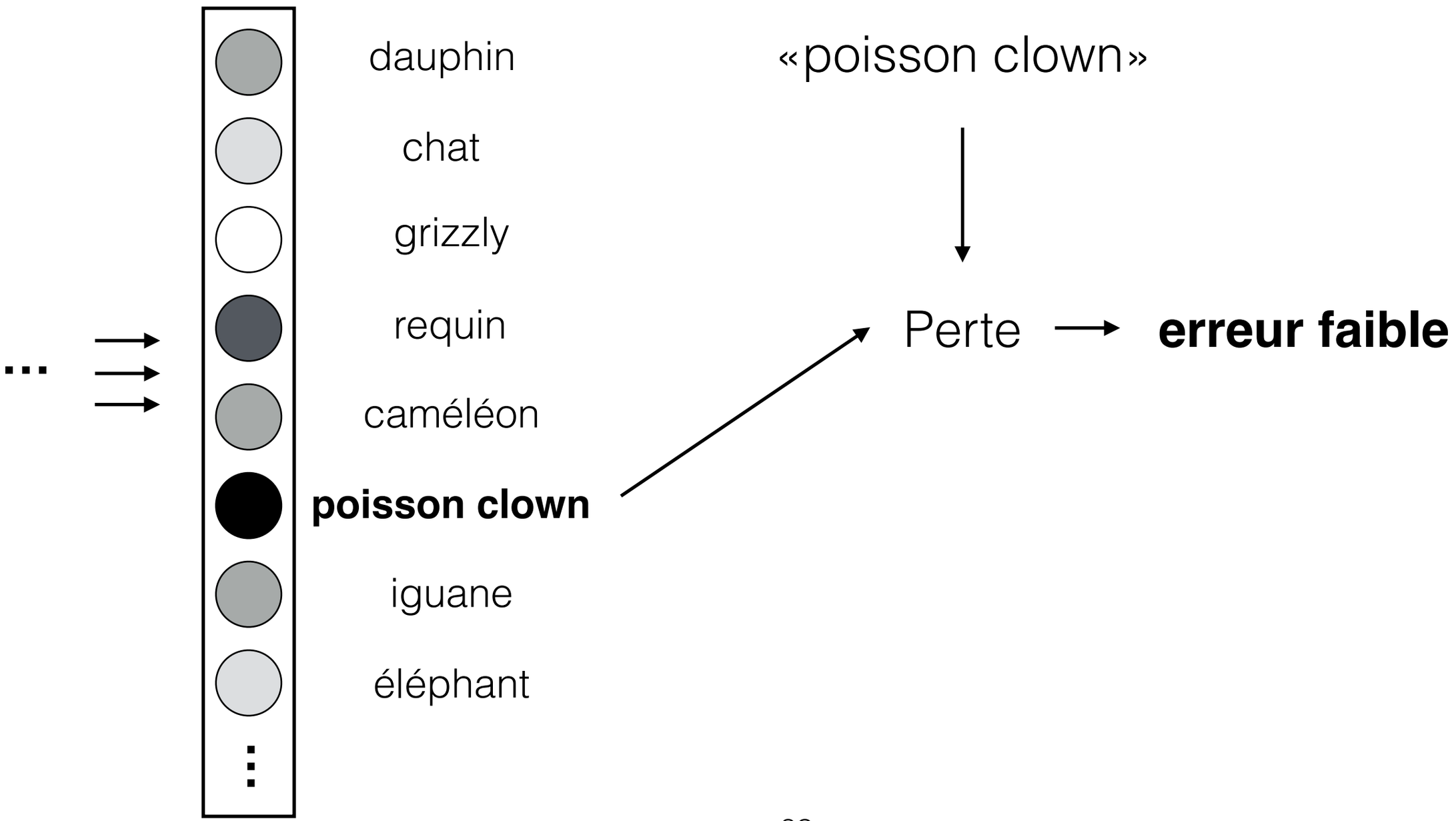
Perte → $L()$

$$\underset{\mathbf{w}}{\text{argmin}} \quad L(w_1, \ldots, w_6)$$

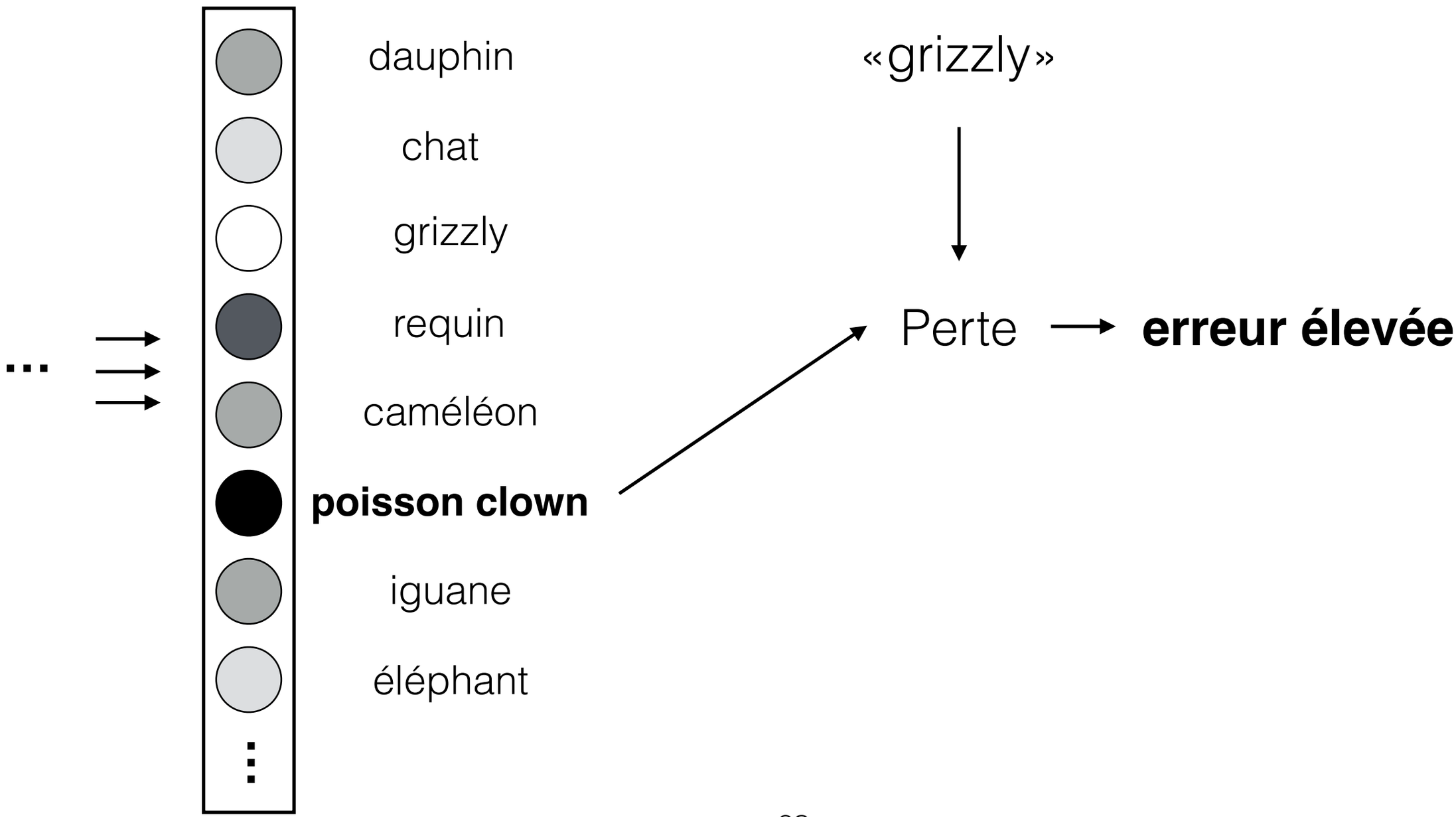
Fonction de perte (*loss function*)



Fonction de perte (*loss function*)

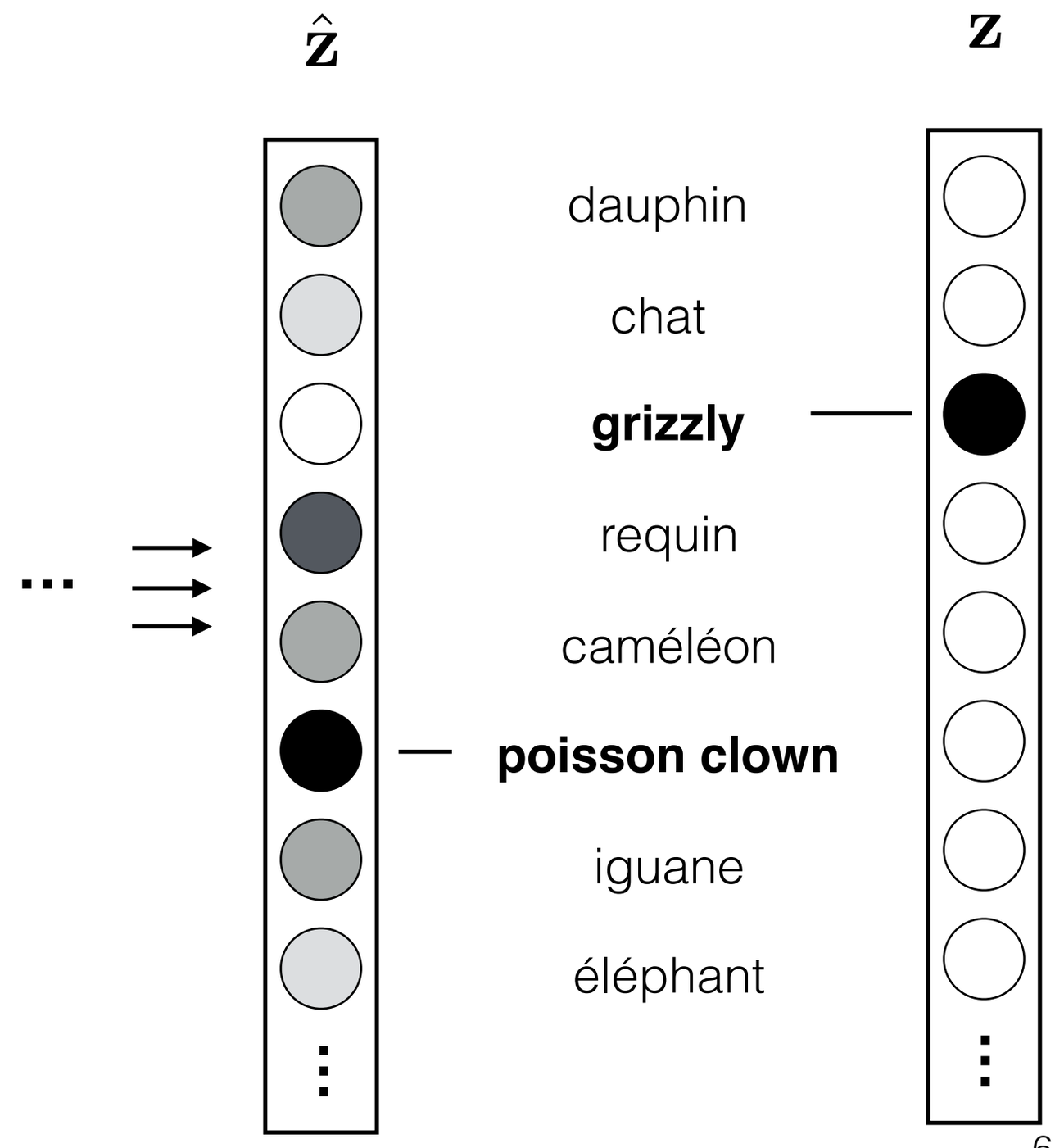


Fonction de perte (*loss function*)



Fonction de perte pour classification

Sortie du réseau «Vraie» étiquette



Probabilité des données observées sous le modèle

$$H(\hat{\mathbf{z}}, \mathbf{z}) = -\sum_c \hat{\mathbf{z}}_c \log \mathbf{z}_c$$

Apprend un modèle de type $p(c|\mathbf{x})$

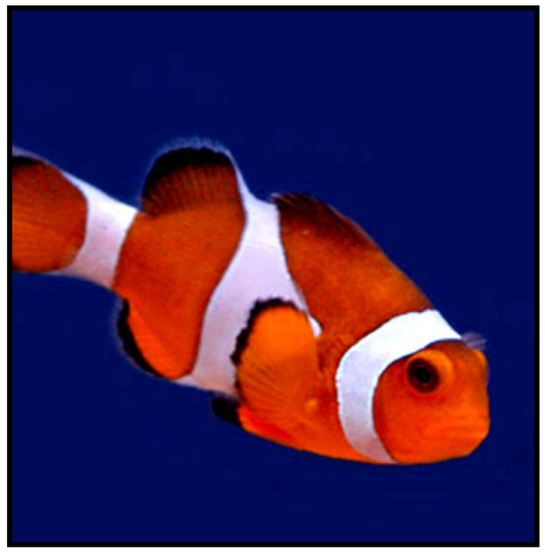
Apprentissage par réseaux profonds

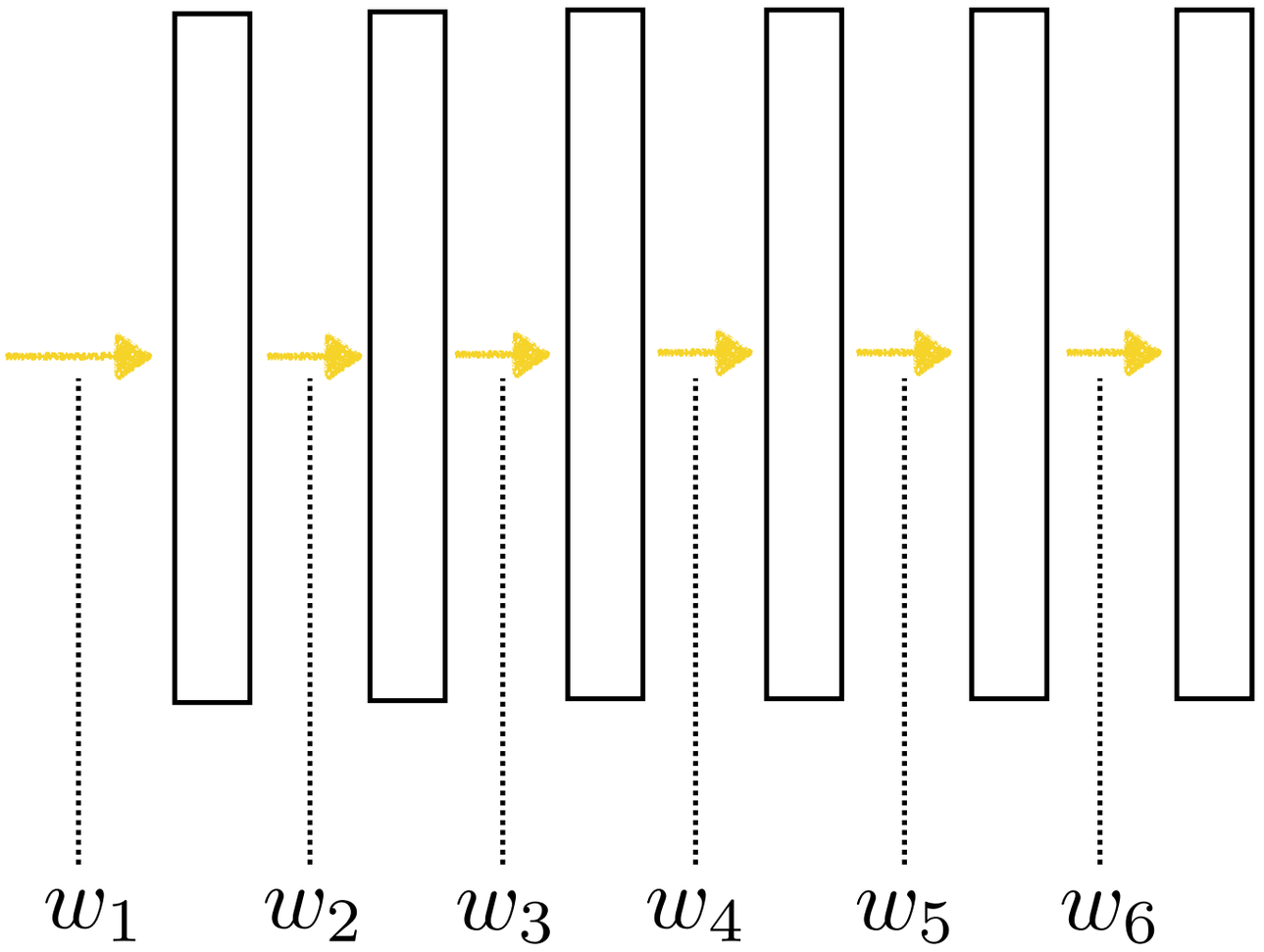
Appris automatiquement

$\mathbf{z}_1$

«poisson clown»

$\mathbf{x}_1$





Perte

$\ell(\mathbf{z}_1, f(\mathbf{x}_1; \mathbf{w}))$

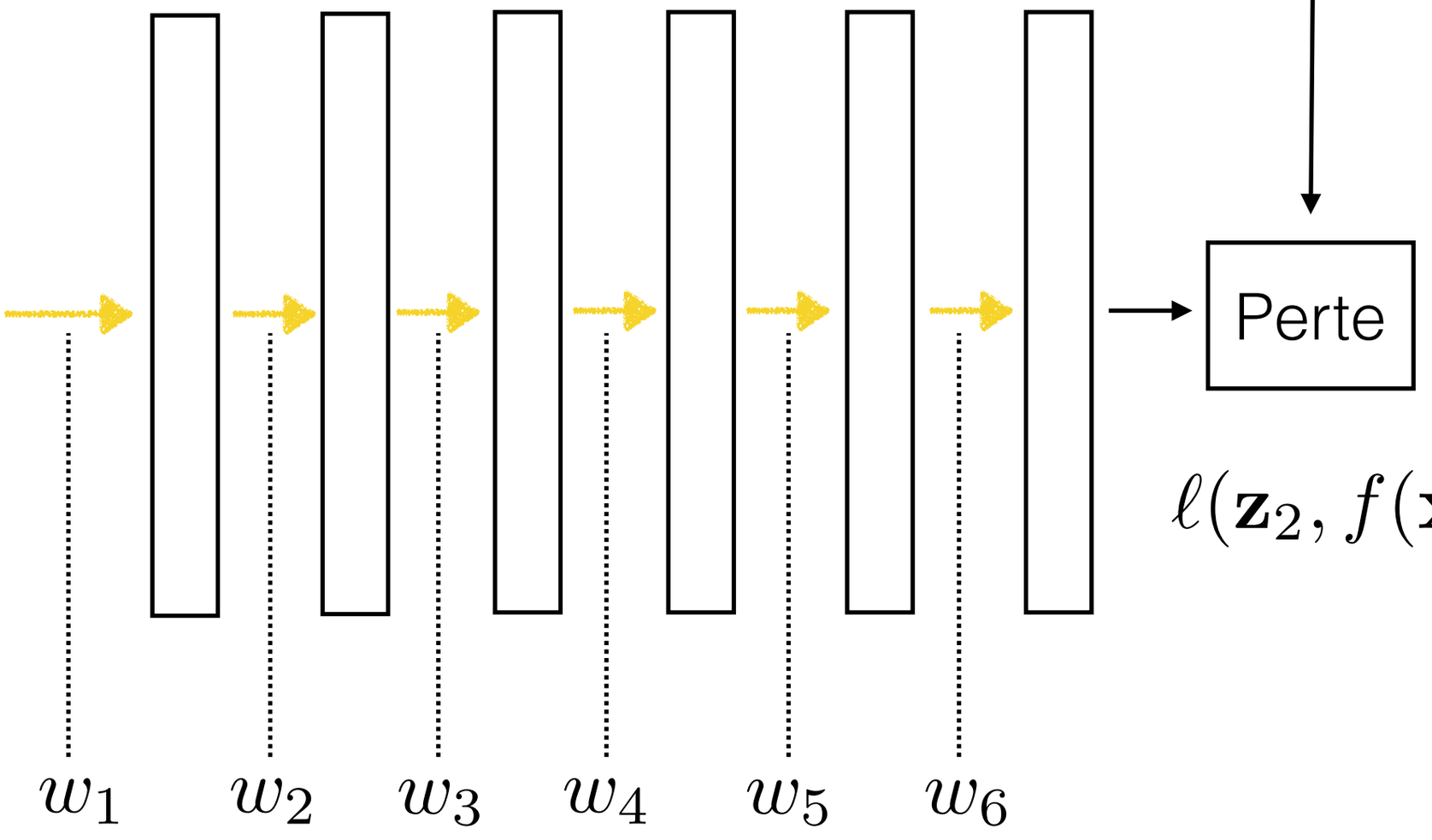
Apprentissage par réseaux profonds

Appris automatiquement

$\mathbf{z}_2$

«grizzly»

$\mathbf{x}_2$



$\ell(\mathbf{z}_2, f(\mathbf{x}_2; \mathbf{w}))$

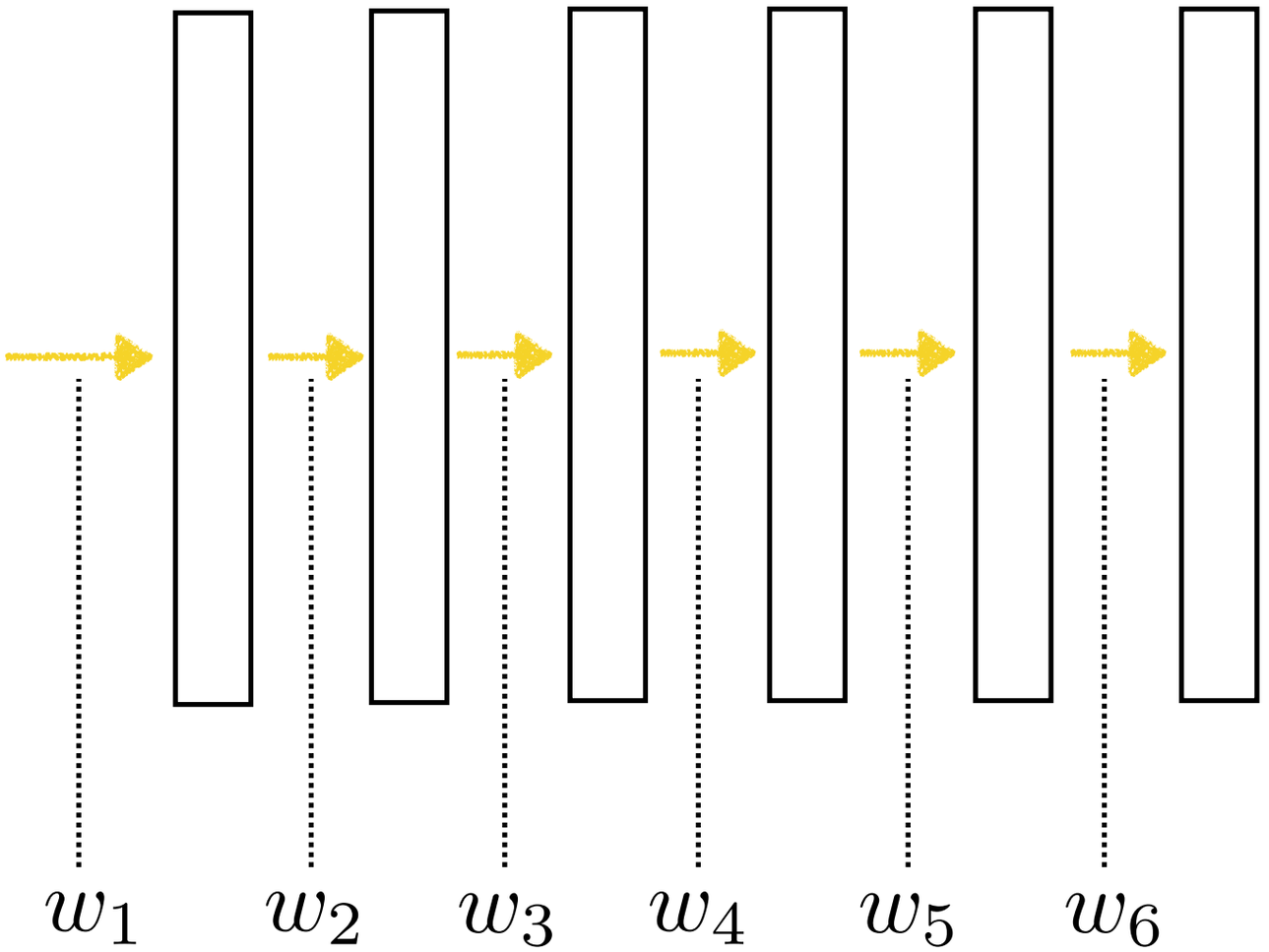
Apprentissage par réseaux profonds

Appris automatiquement

$\mathbf{z}_i$

«caméléon»

$\mathbf{x}_i$



Perte

$\ell(\mathbf{z}_i, f(\mathbf{x}_i; \mathbf{w}))$

$$\underset{\mathbf{w}}{\operatorname{argmin}} \sum_i \ell(\mathbf{z}_i, f(\mathbf{x}_i; \mathbf{w}))$$

P. Isola

Descente du gradient

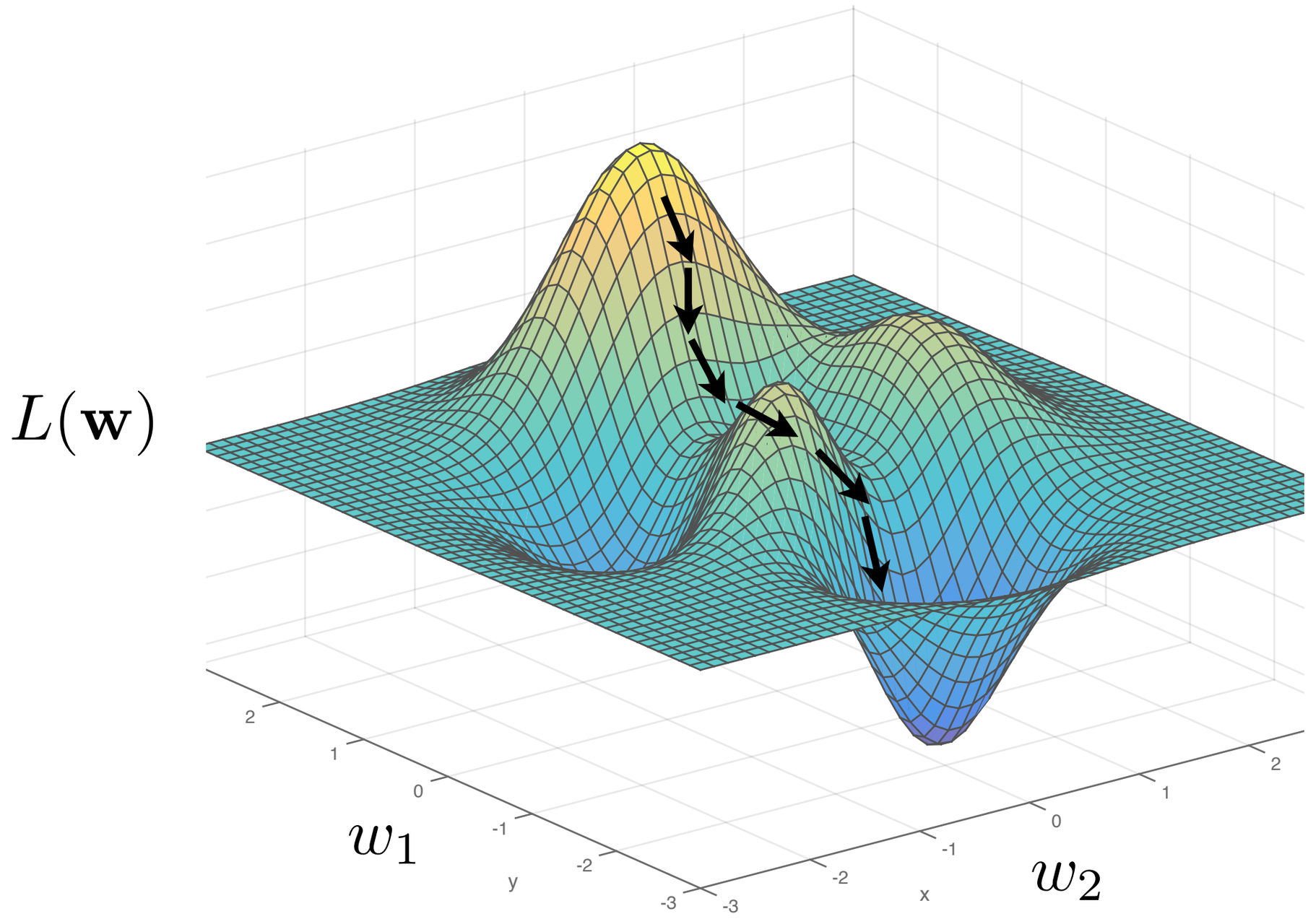
$$\underset{\mathbf{w}}{\operatorname{argmin}} \quad \sum_i \ell(\mathbf{z}_i, f(\mathbf{x}_i; \mathbf{w})) = L(\mathbf{w})$$

Une itération de descente du gradient

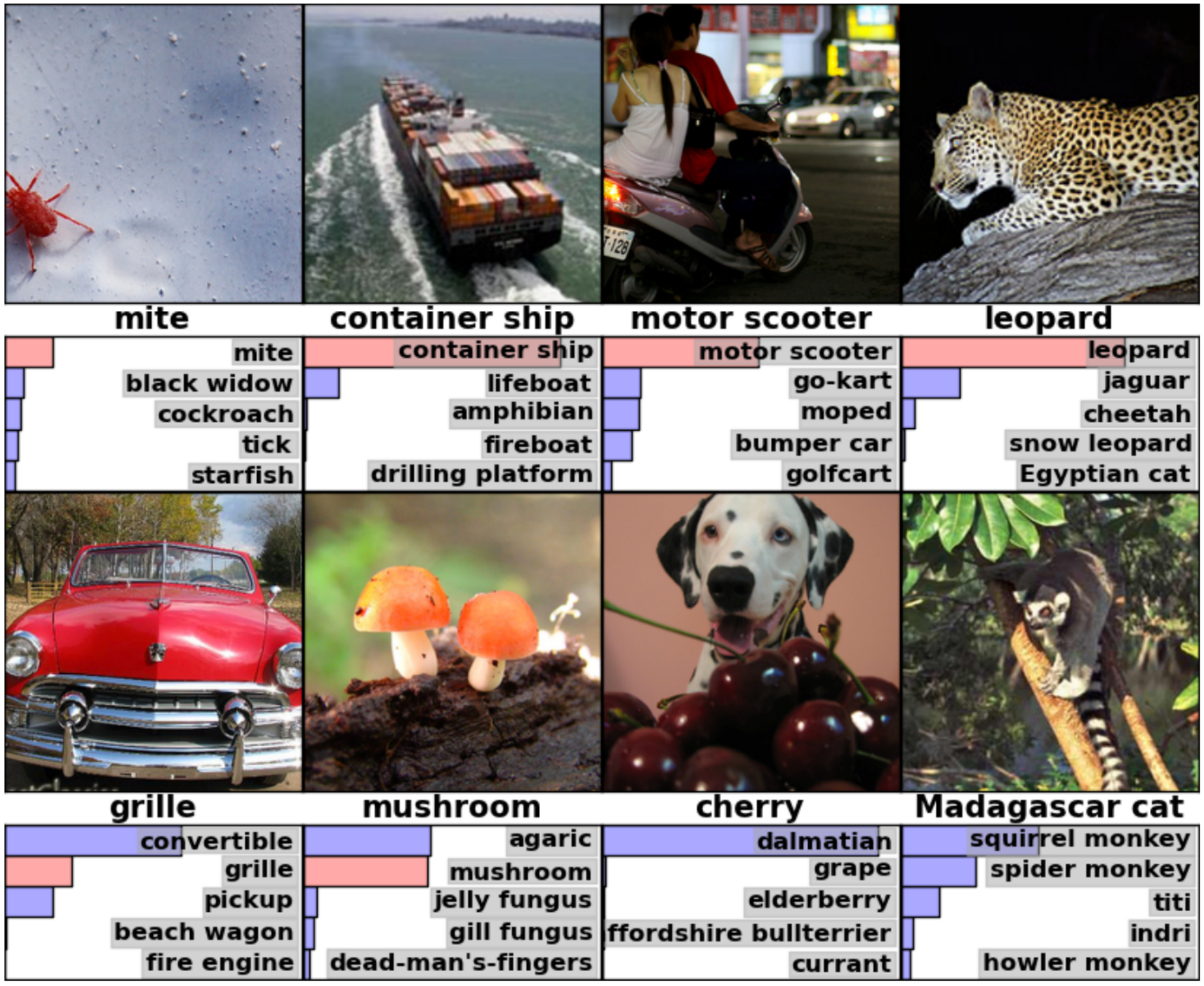
$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta_t \frac{\partial L(\mathbf{w^t})}{\partial \mathbf{w}}$$

taux d'apprentissage

Descente du gradient

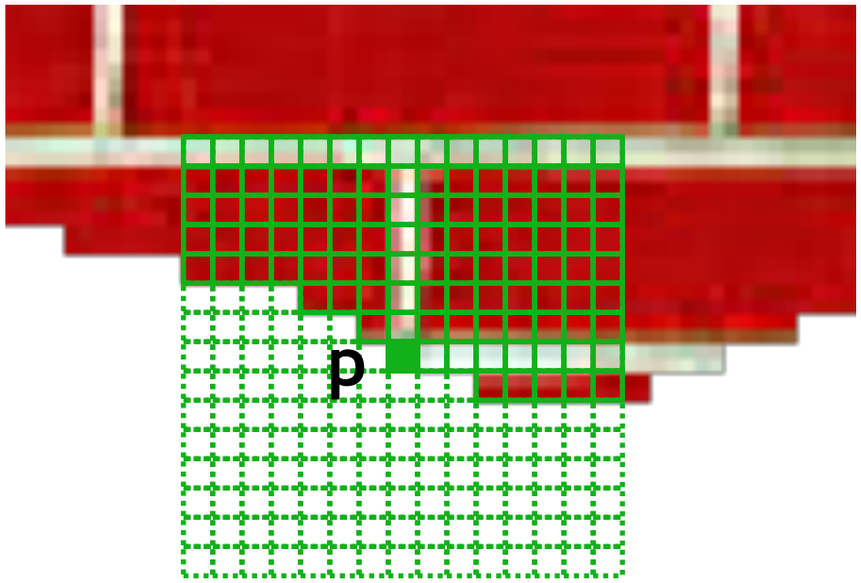


$$p(c|\mathbf{x})$$



[Krizhevsky et al. NIPS 2012]

Synthèse de texture: rappel



Synthétise un pixel

échantillonnage
non-paramétrique

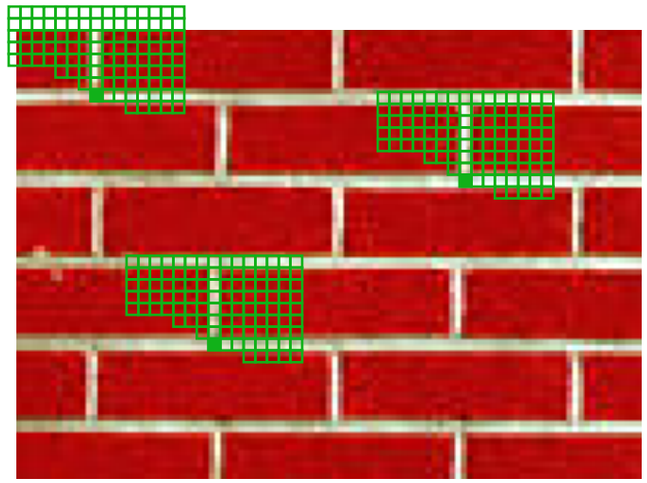
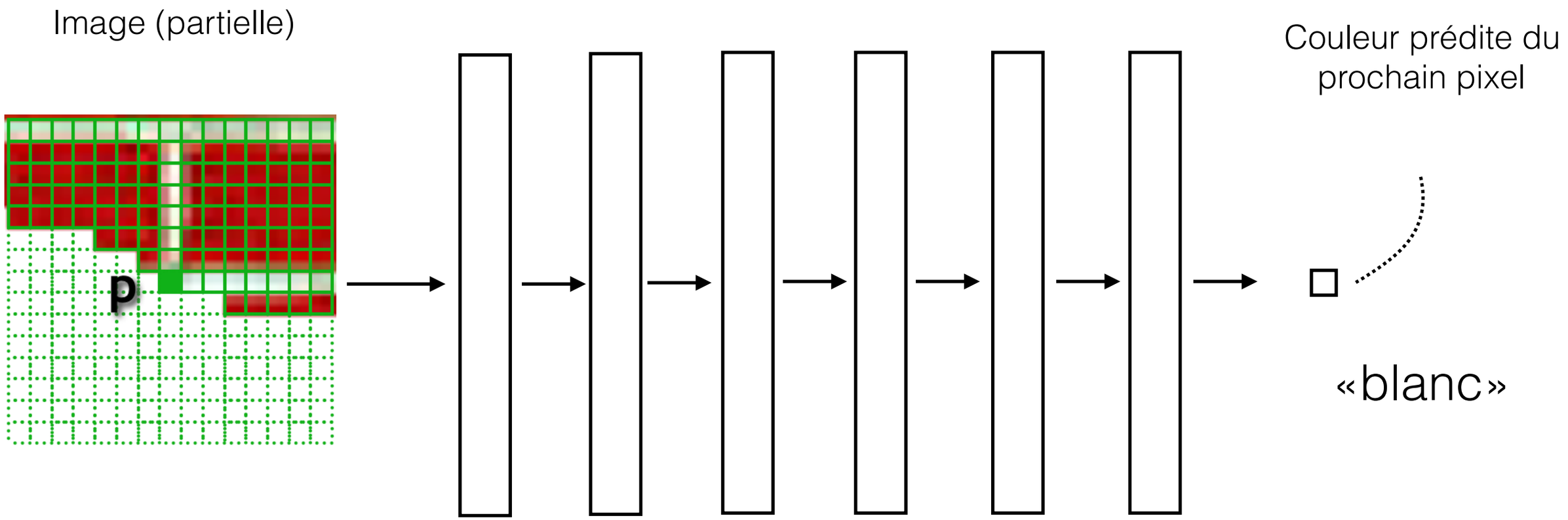


Image d'entrée

Modélise $P(p|N(p))$

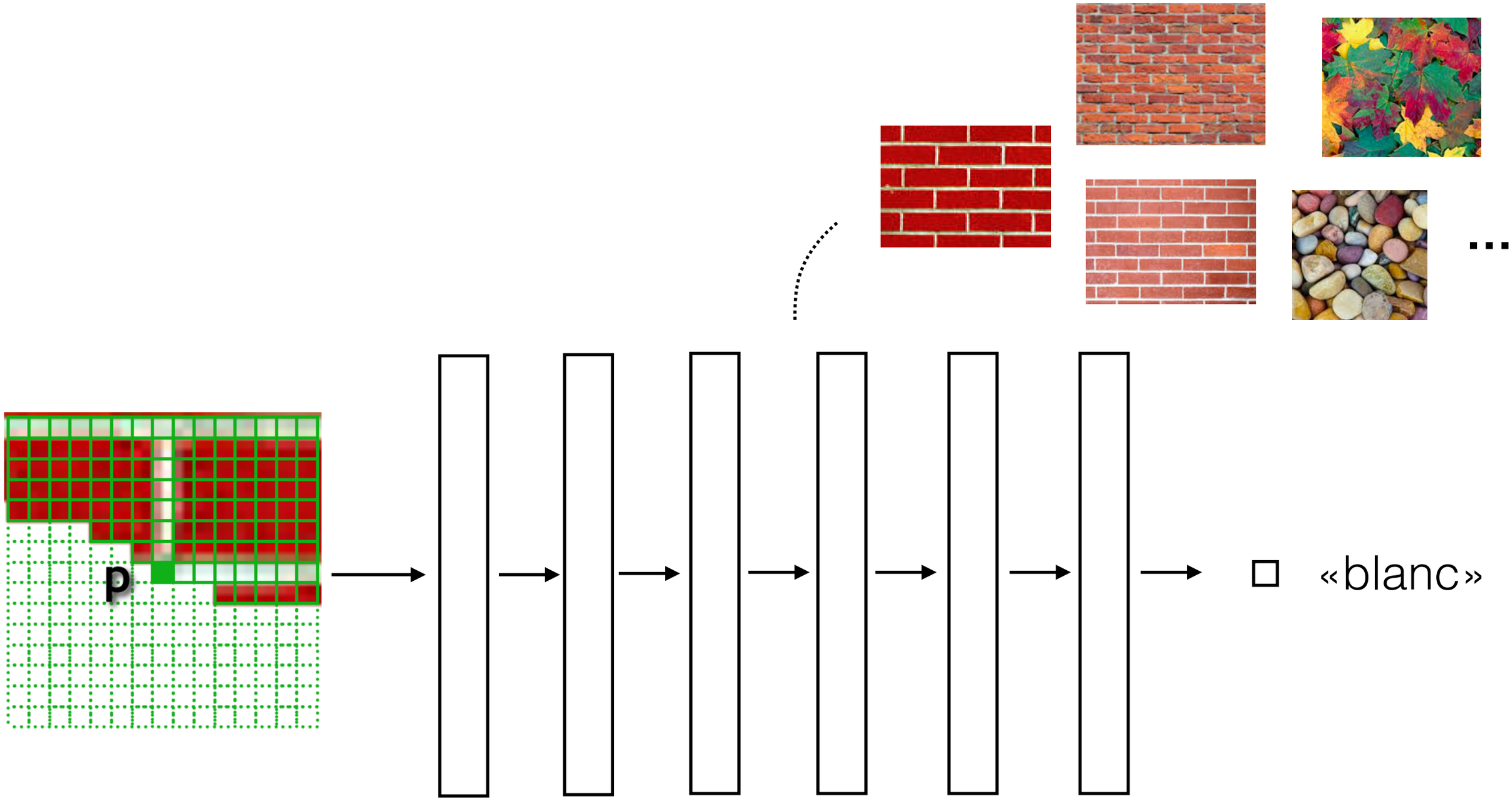
[Efros & Leung 1999]

Synthèse de texture par réseau profond



[van der Oord et al. 2016]

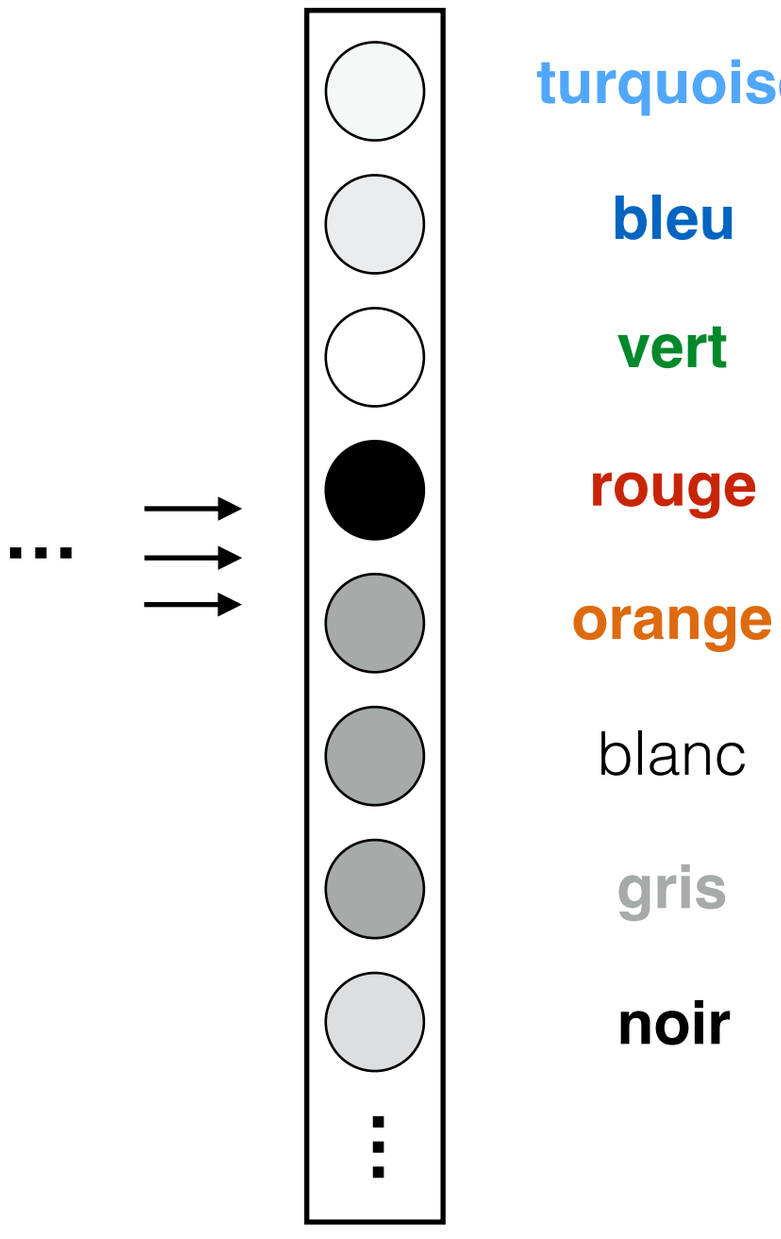
Synthèse de texture par réseau profond



[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau



turquoise

bleu

vert

rouge

orange

blanc

gris

noir

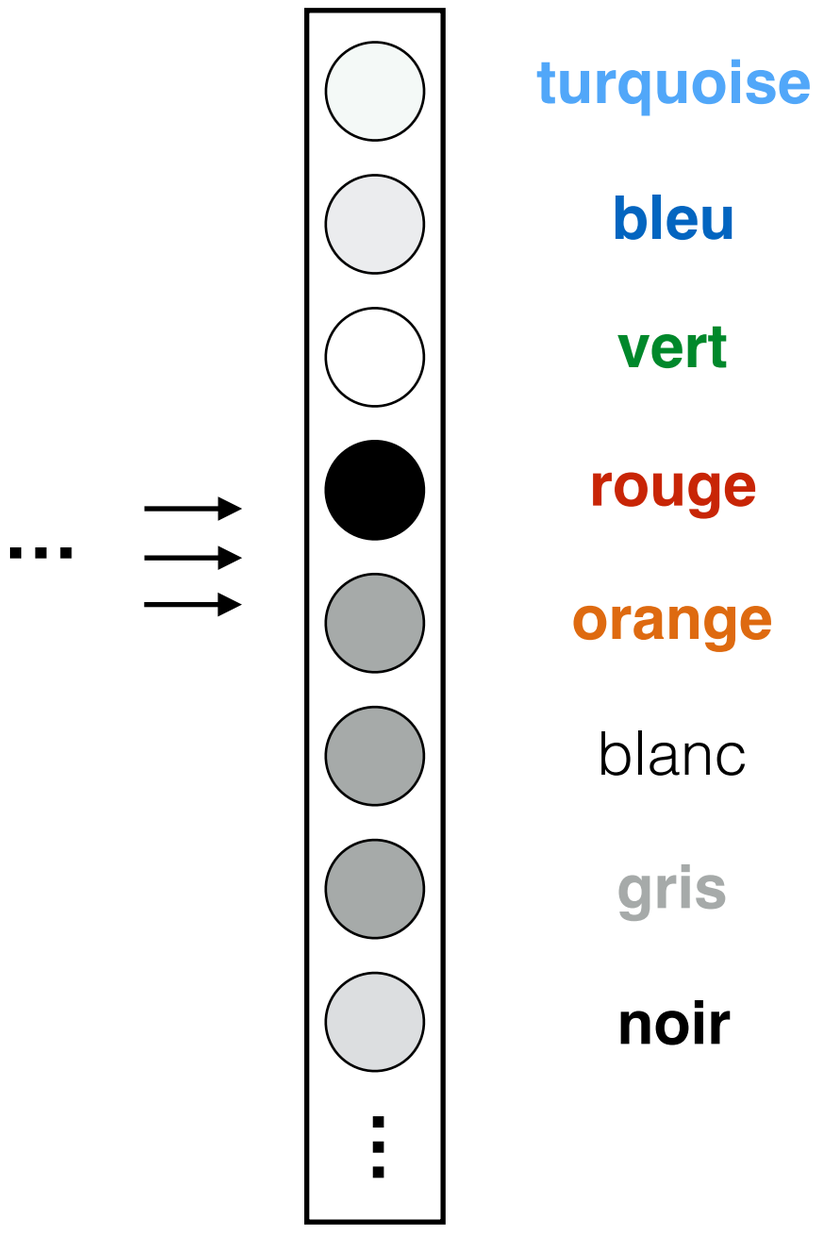
p

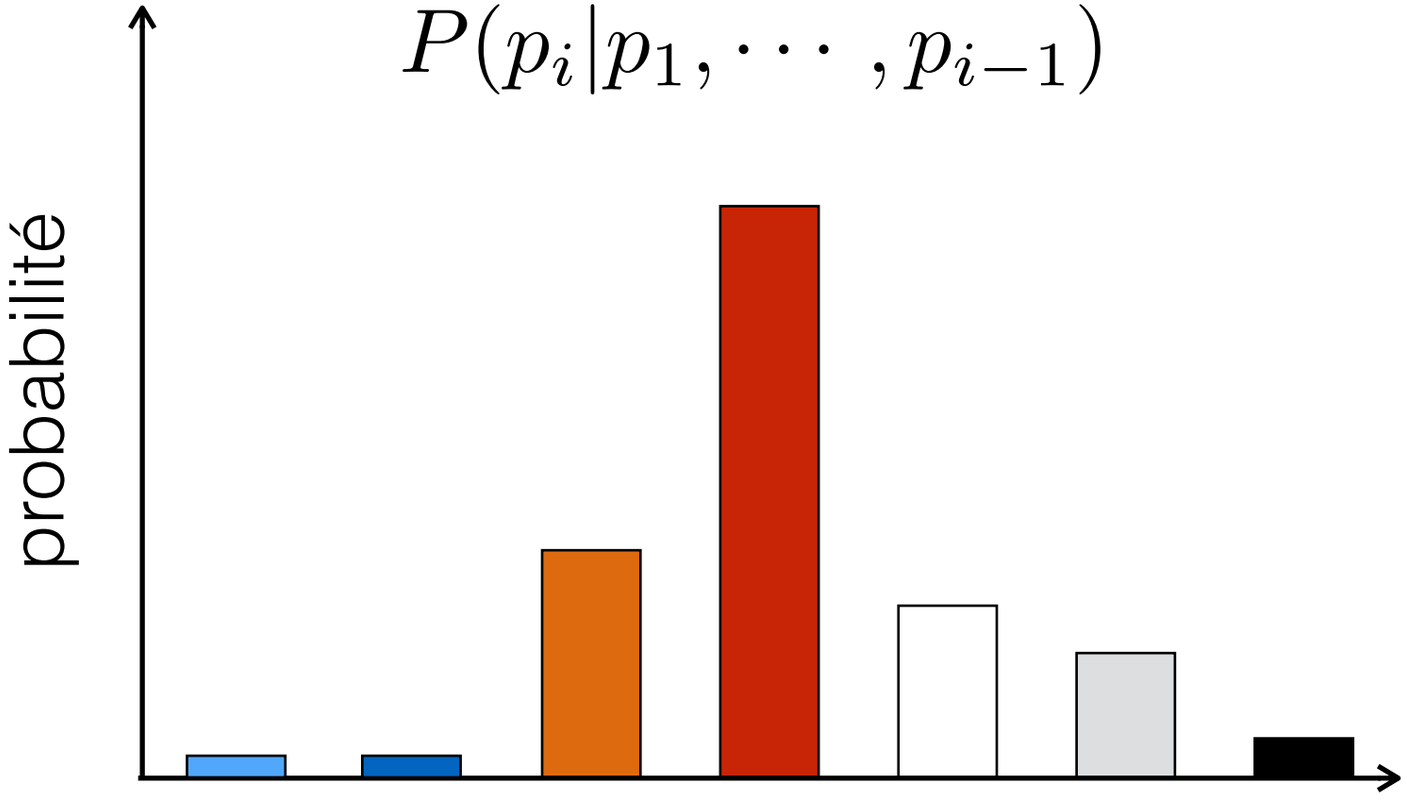
$$P(p_i|p_1, \cdots, p_{i-1})$$

[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau

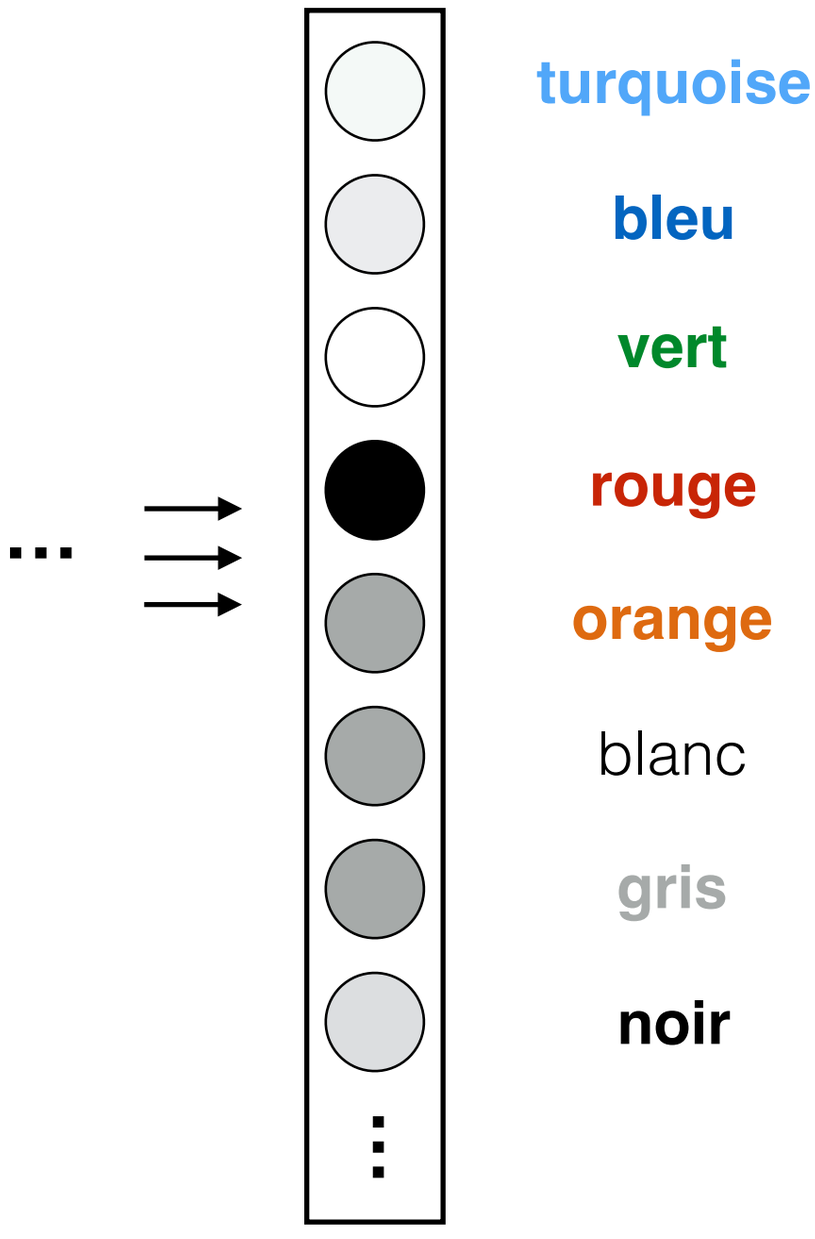


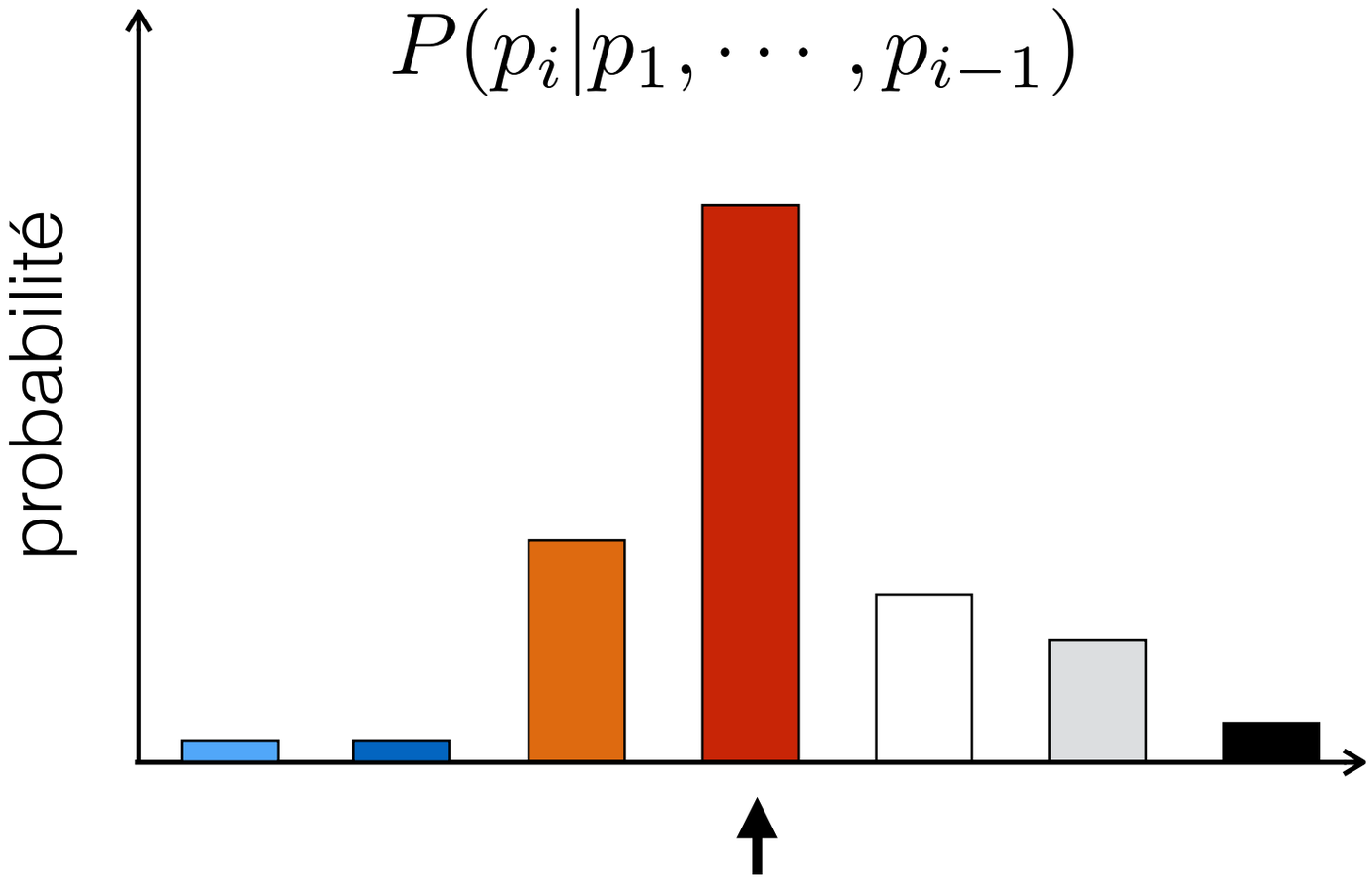


[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau

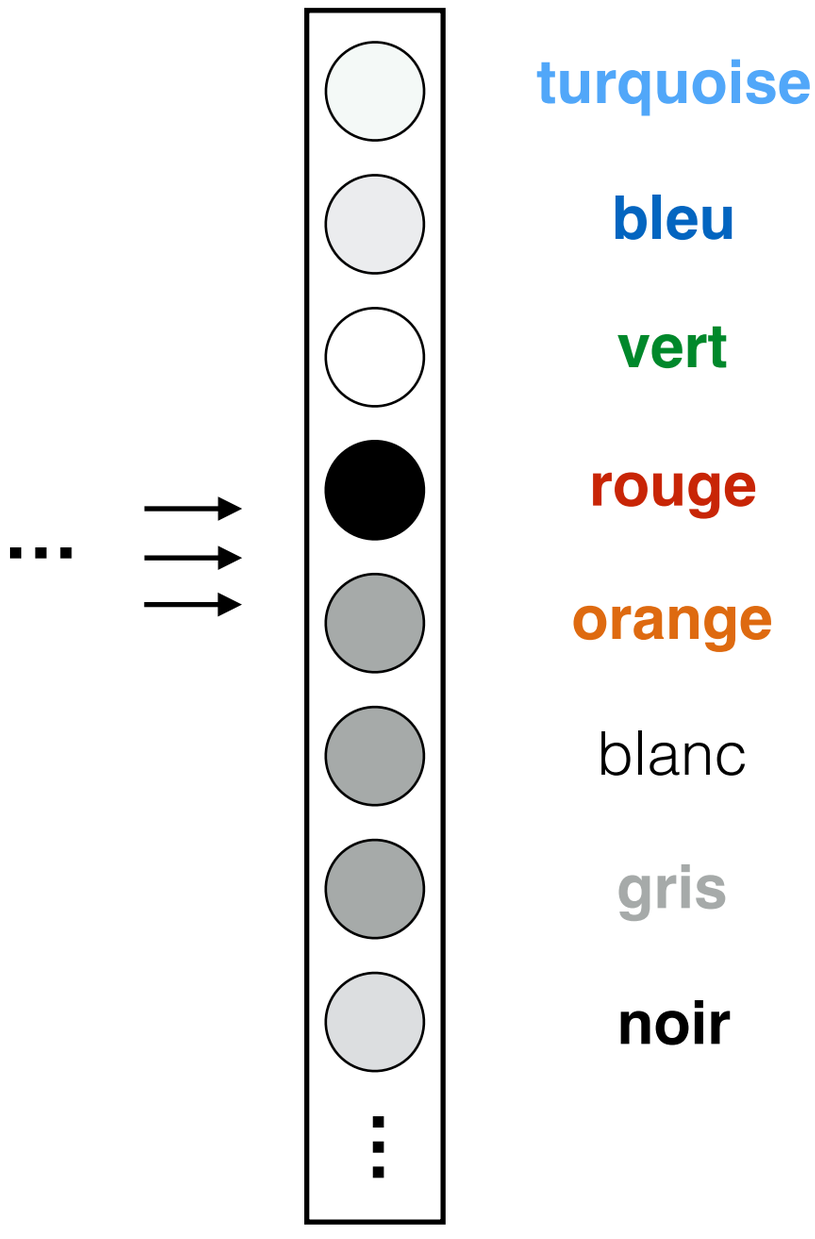


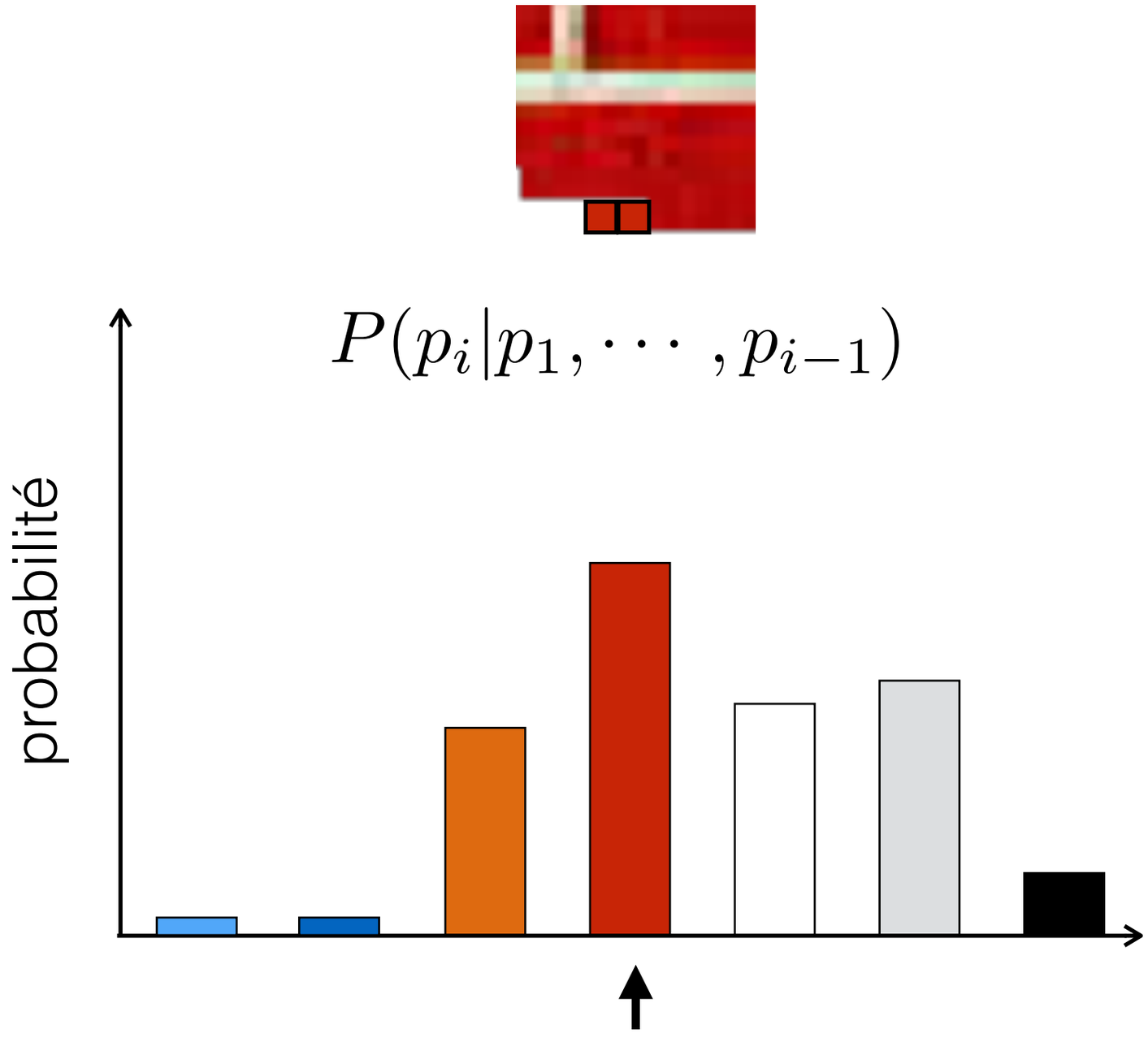


[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau

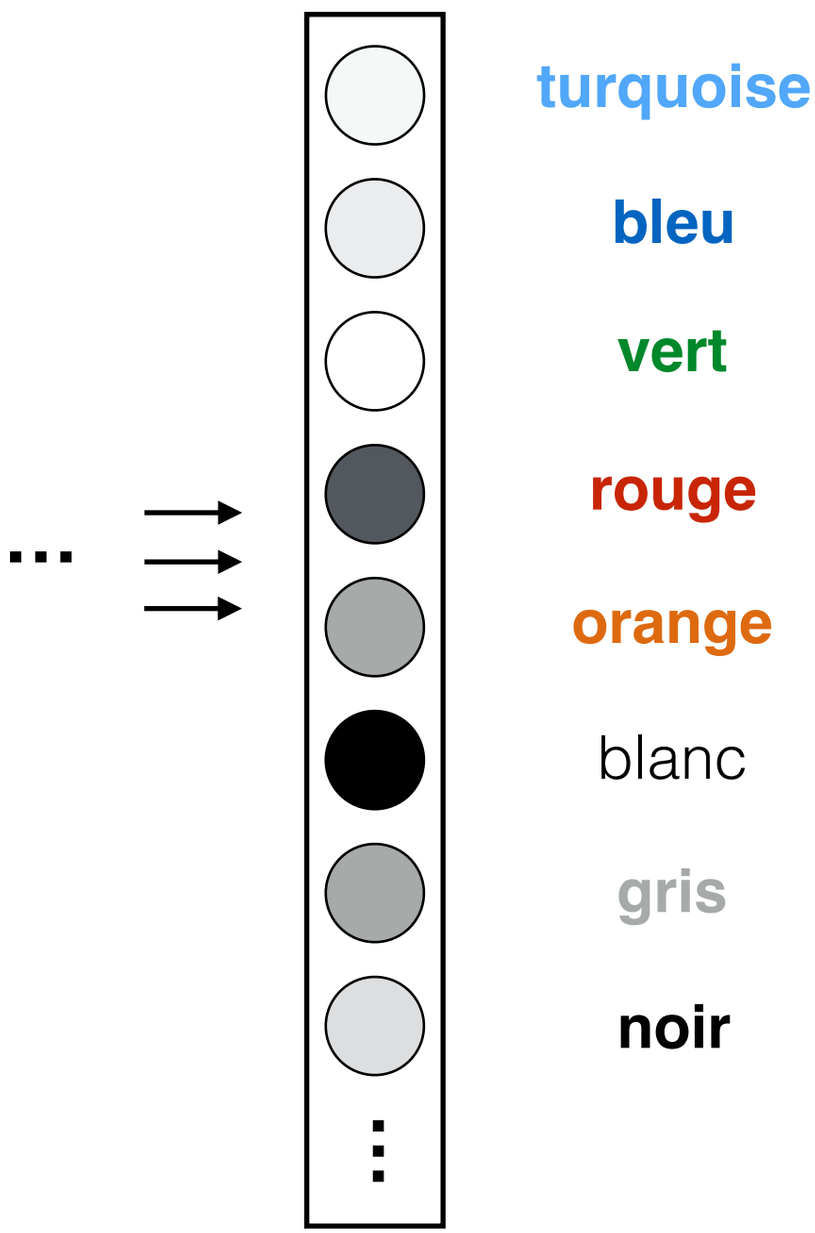


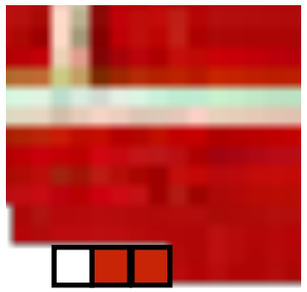


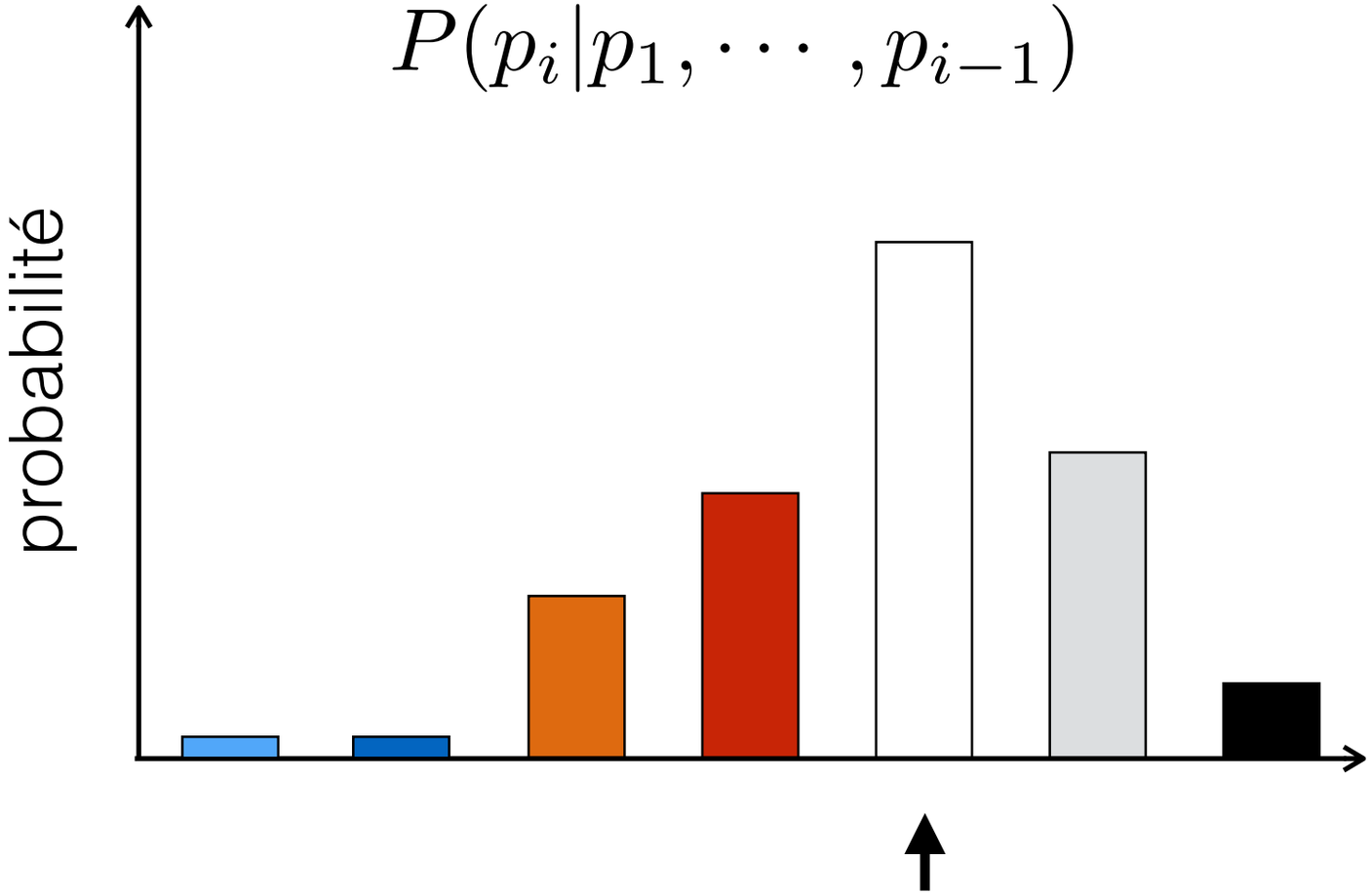
[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau



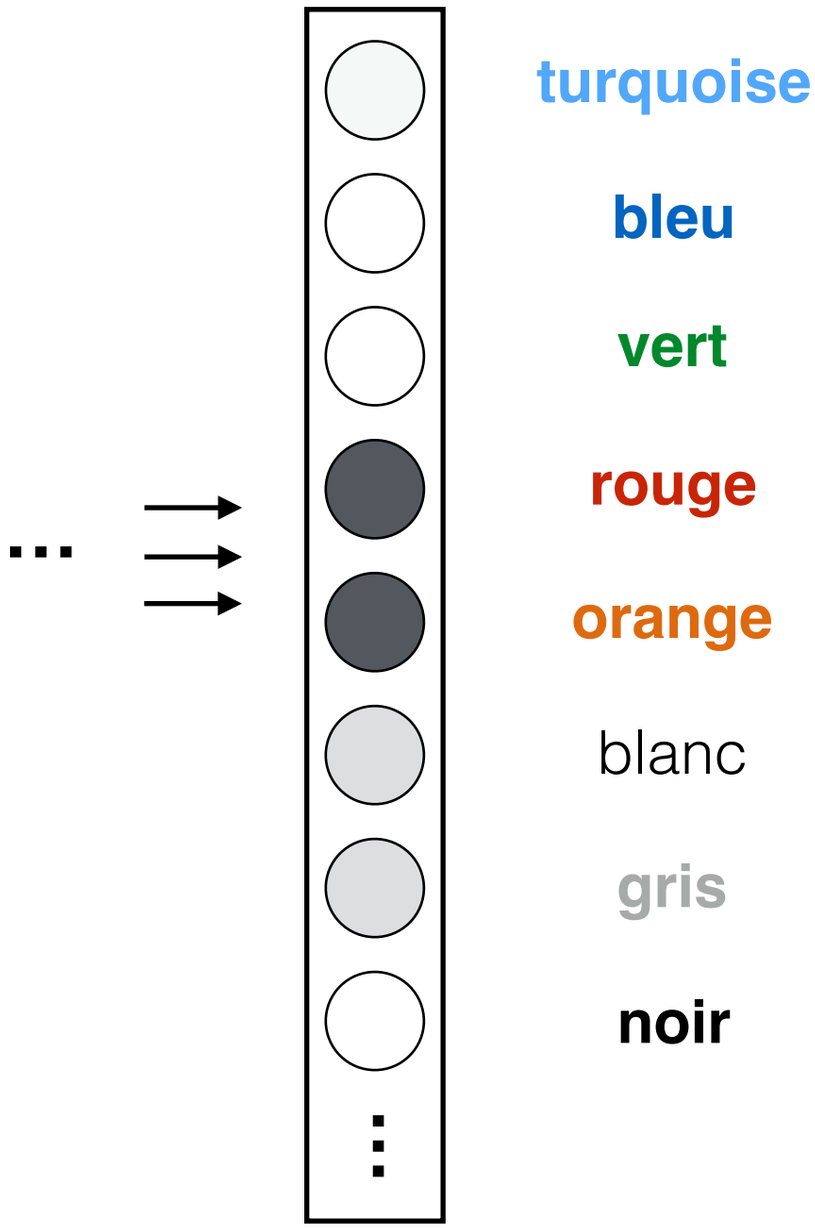


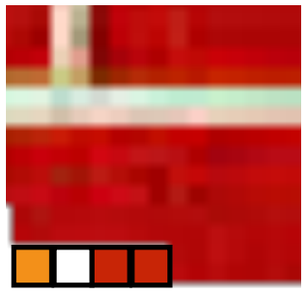


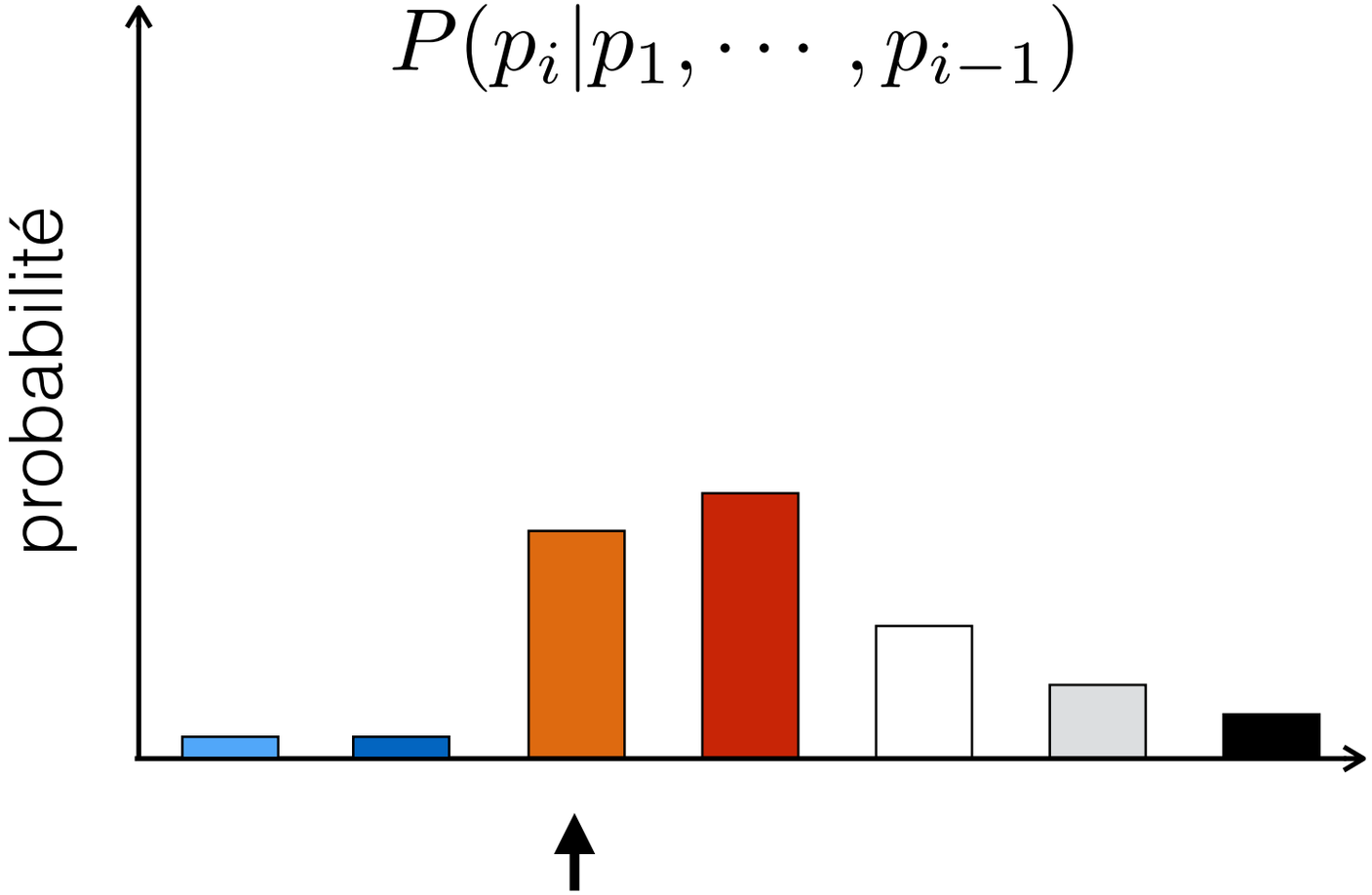
[van der Oord et al. 2016]

Échantillonnage

Sortie du réseau

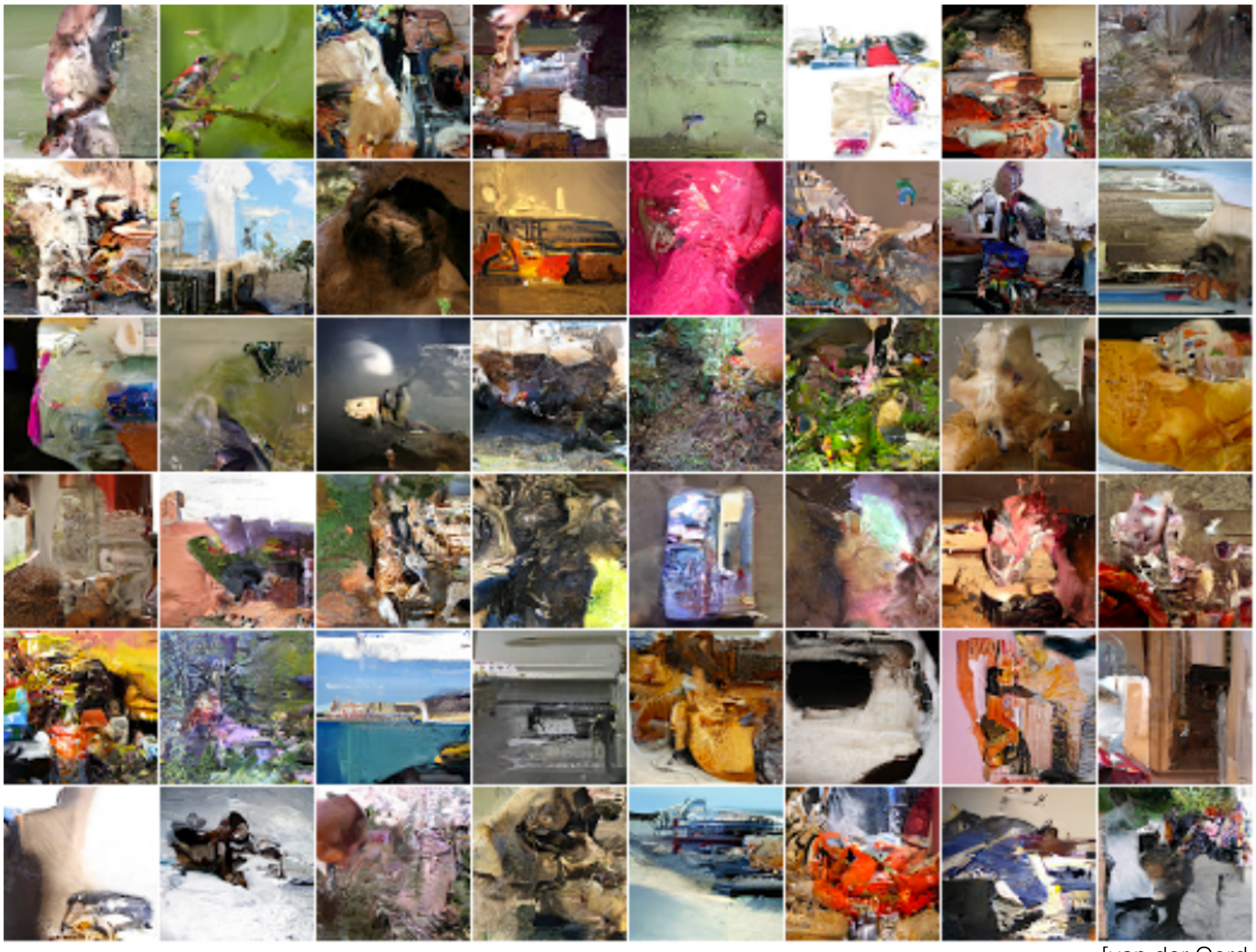






[van der Oord et al. 2016]

Générer des images entières



[van der Oord et al. 2016]

Compléter une image

occlusion

résultats

image originale

[van der Oord et al. 2016]

Ansel Adams, Yosemite Valley Bridge

[Zhang et al. 2016]

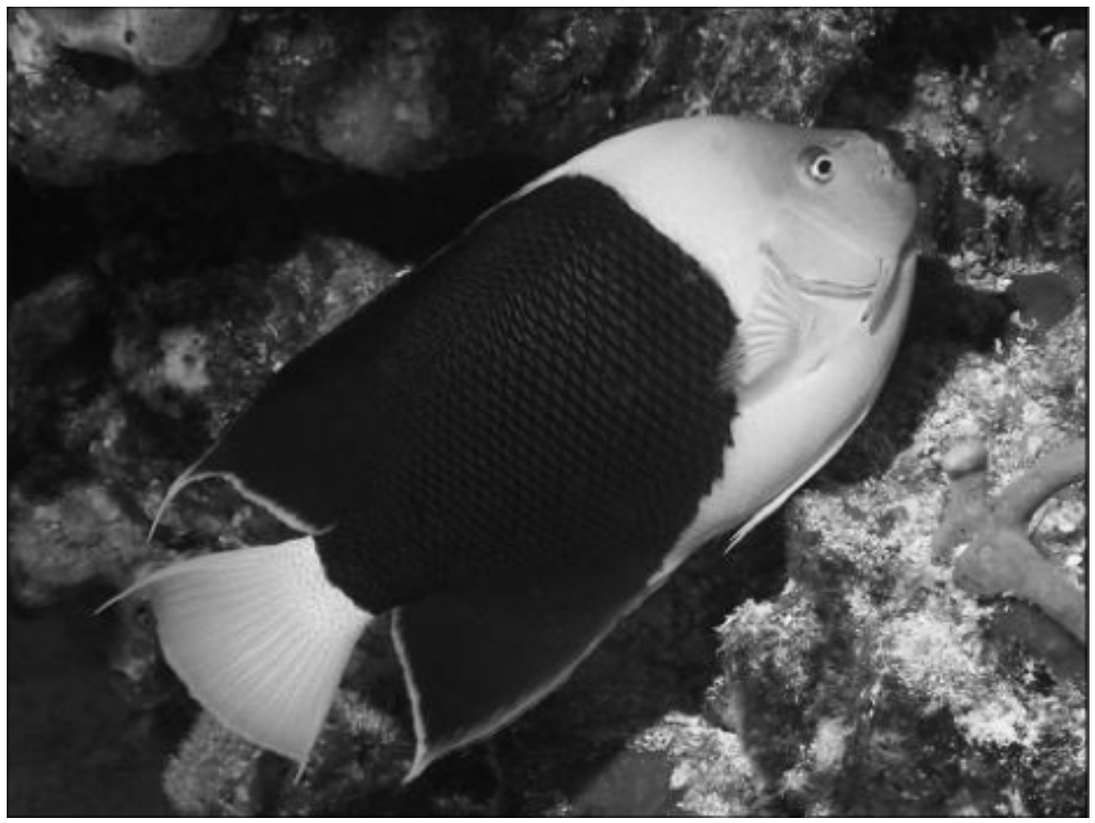


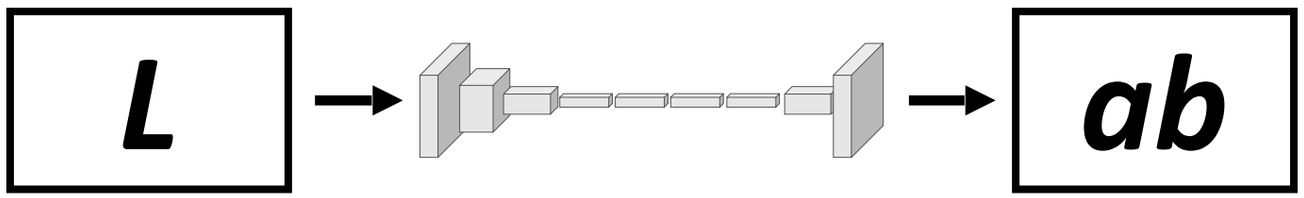
Image d’entrée: canal L seulement

$$\mathcal{F}$$

Information en couleurs: canaux ab (de Lab)

$$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$$

$$\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$$



P. Isola

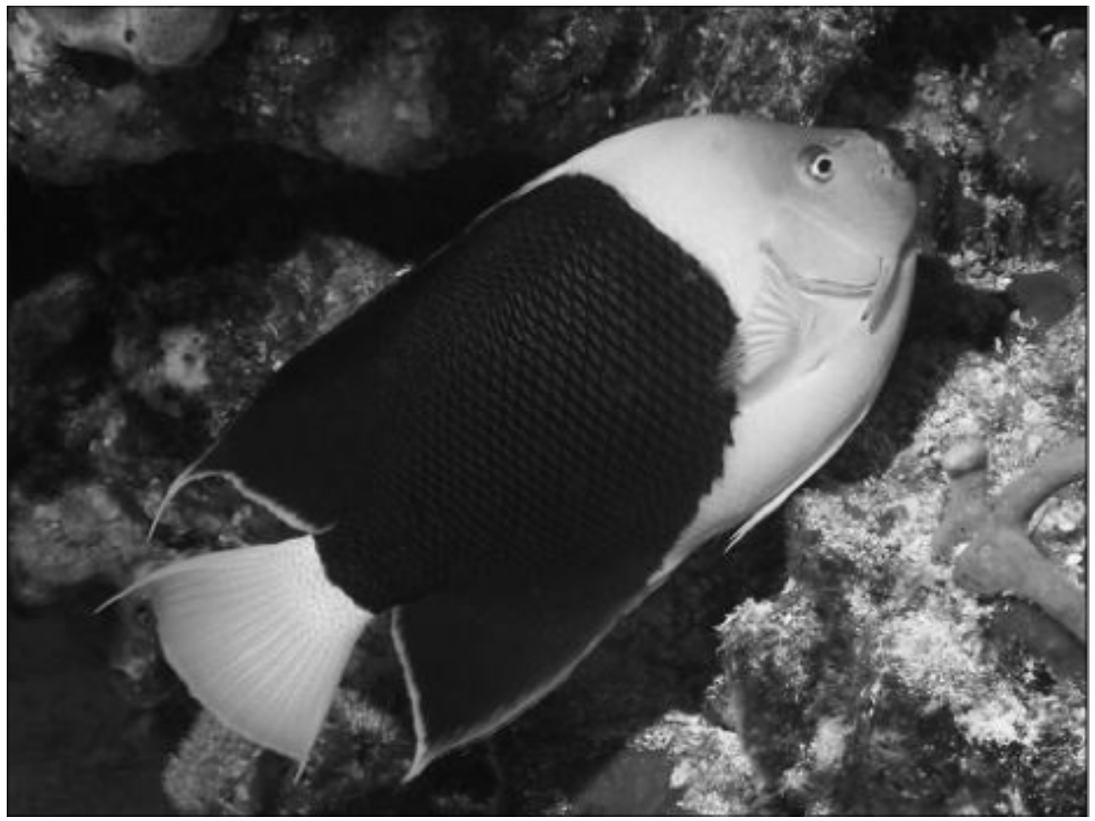
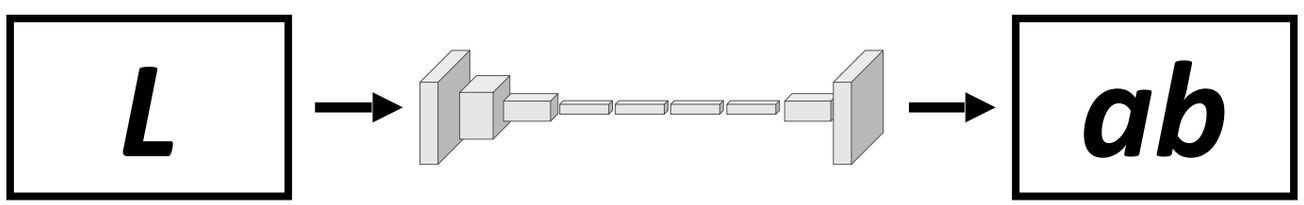


Image: canal L (de Lab)

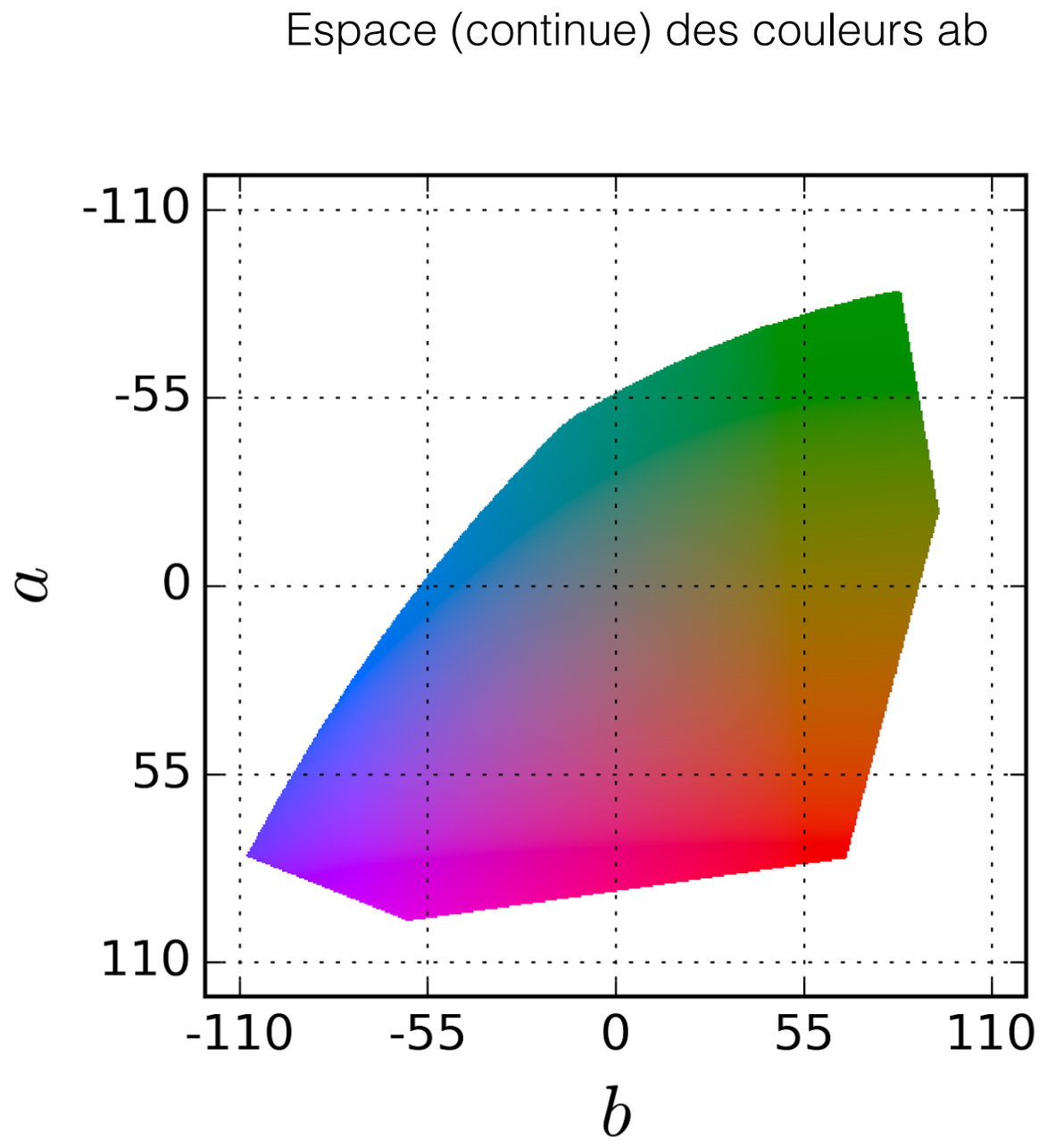
Concatenation (L,ab)

$$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$$

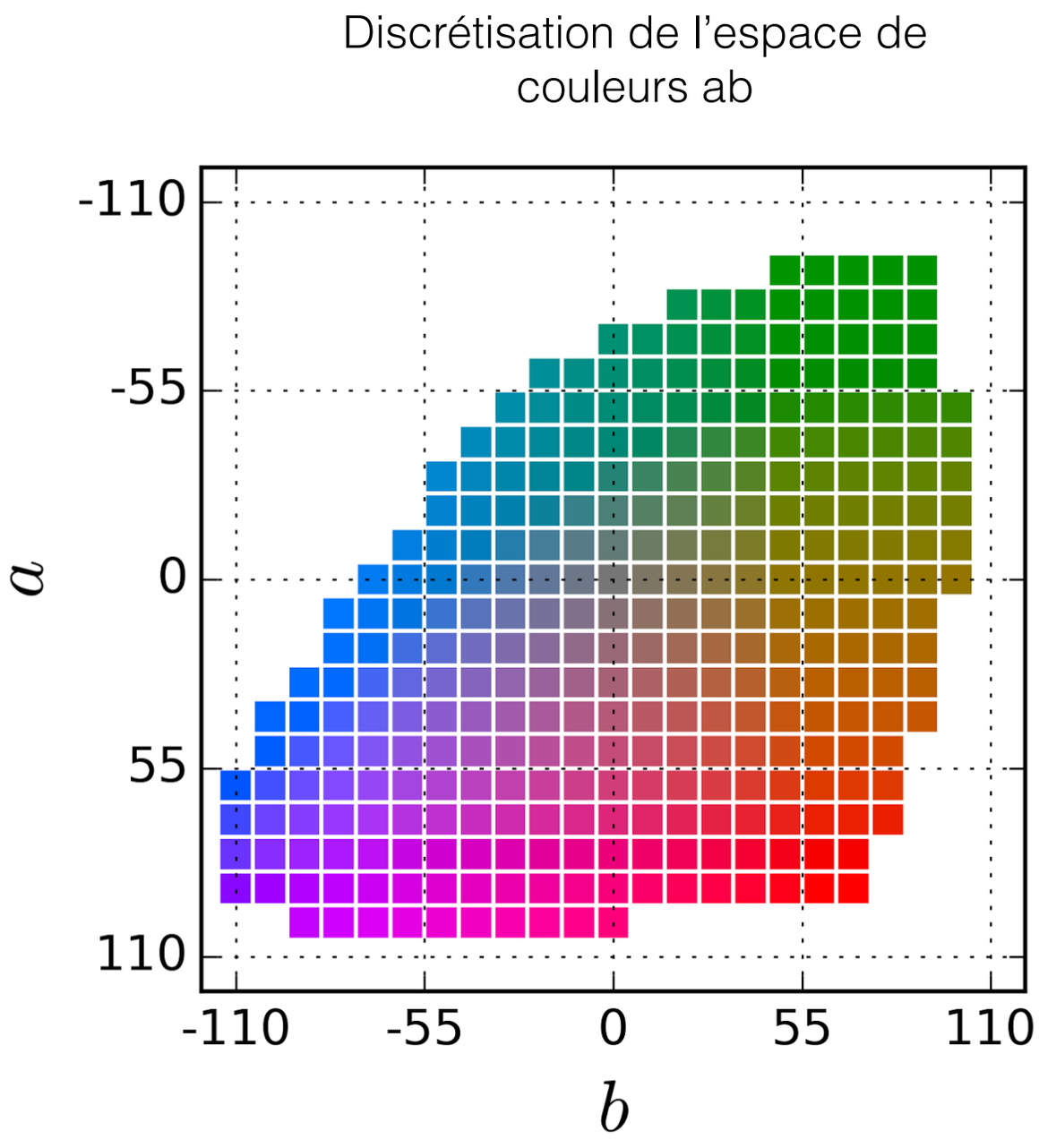
$$(\mathbf{X}, \hat{\mathbf{Y}})$$

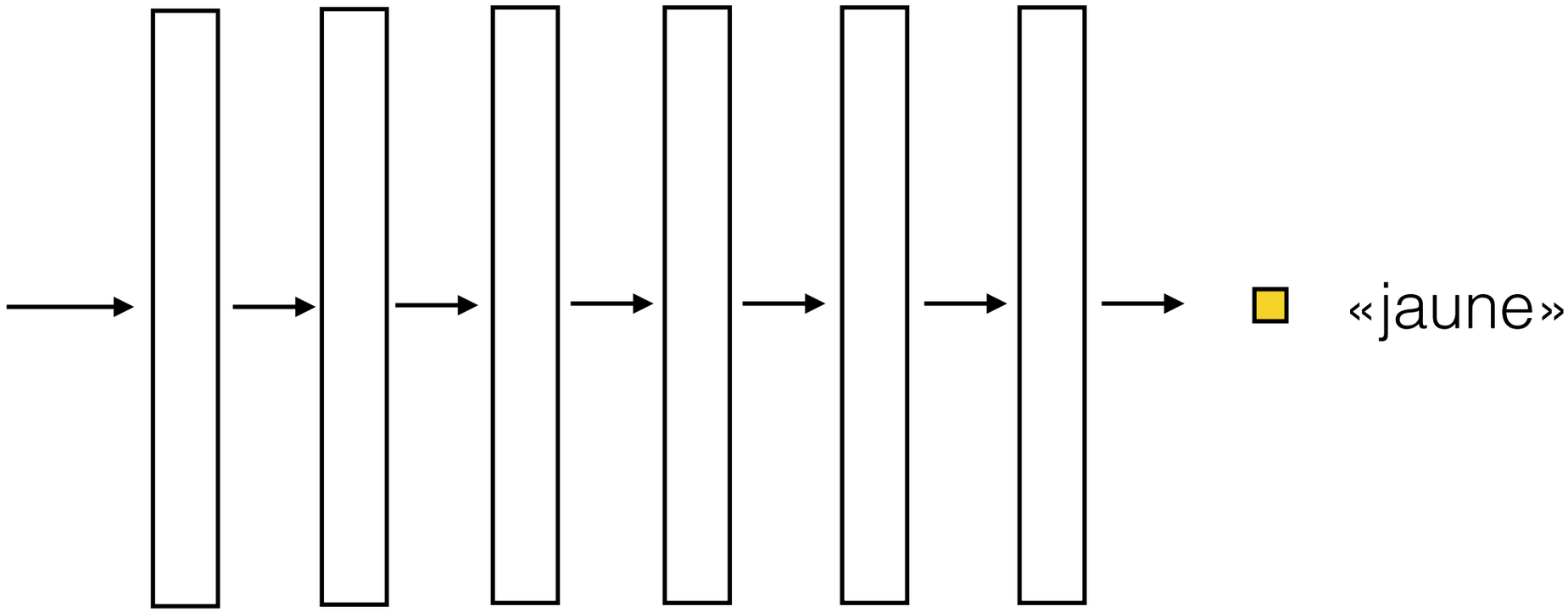


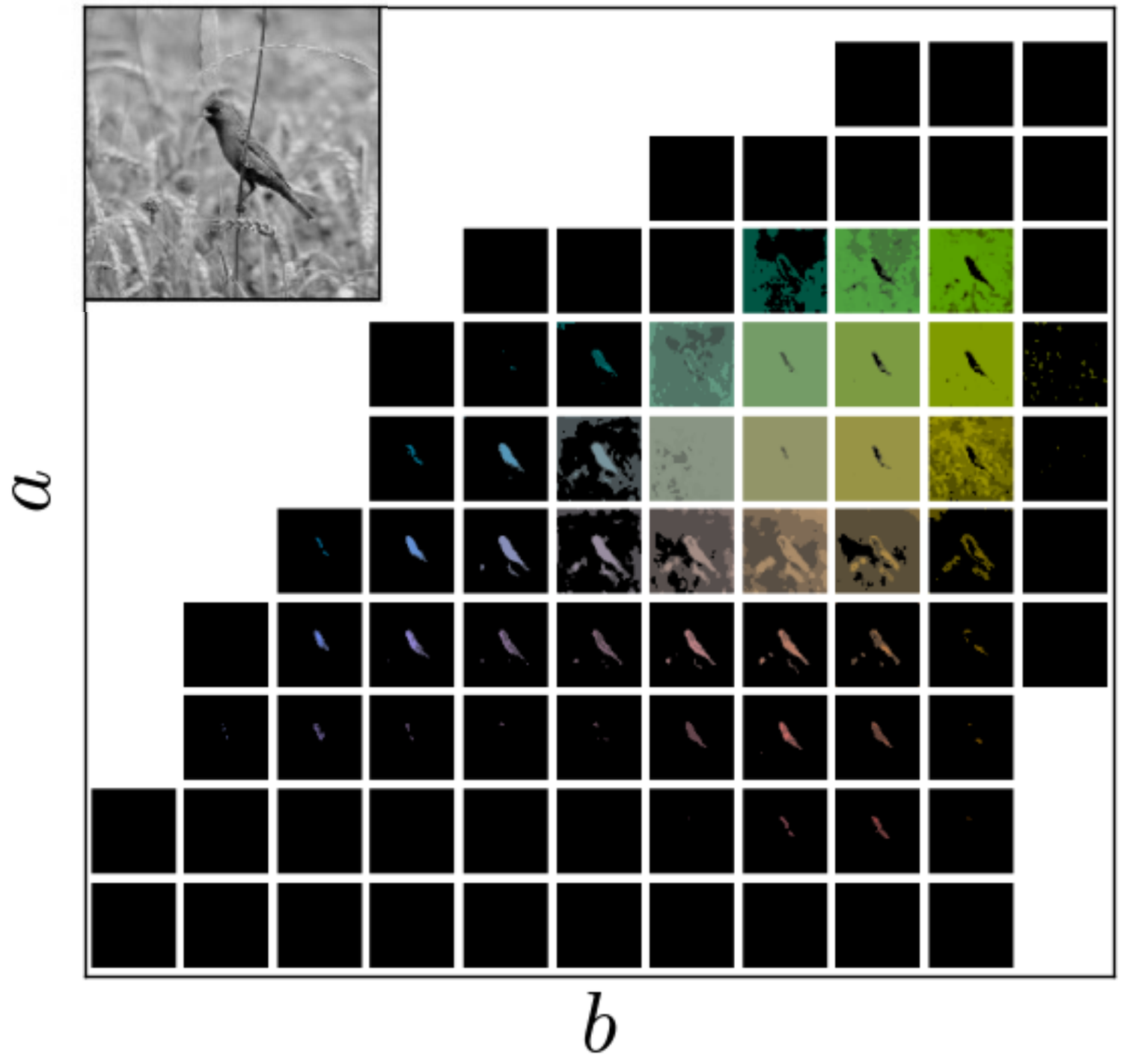
Meilleure fonction de perte



Meilleure fonction de perte



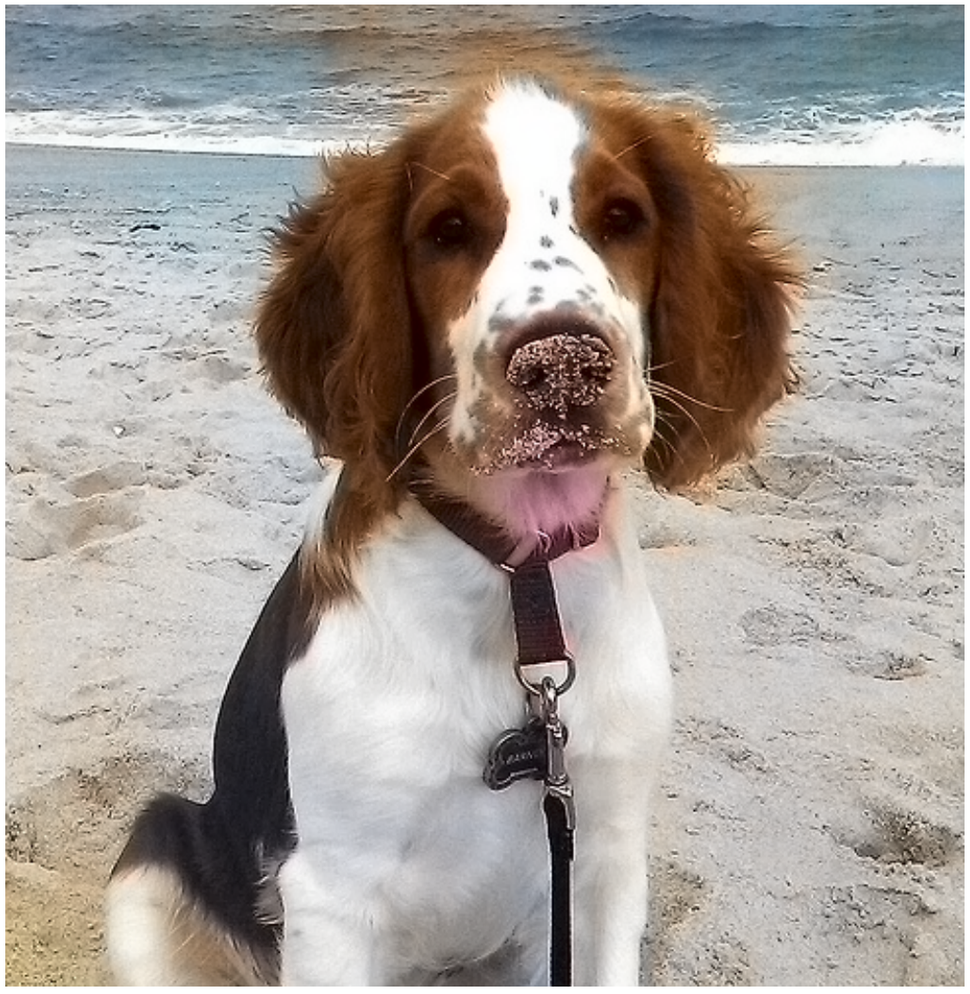

«jaune»


a
b

Bons résultats

Moins bons résultats

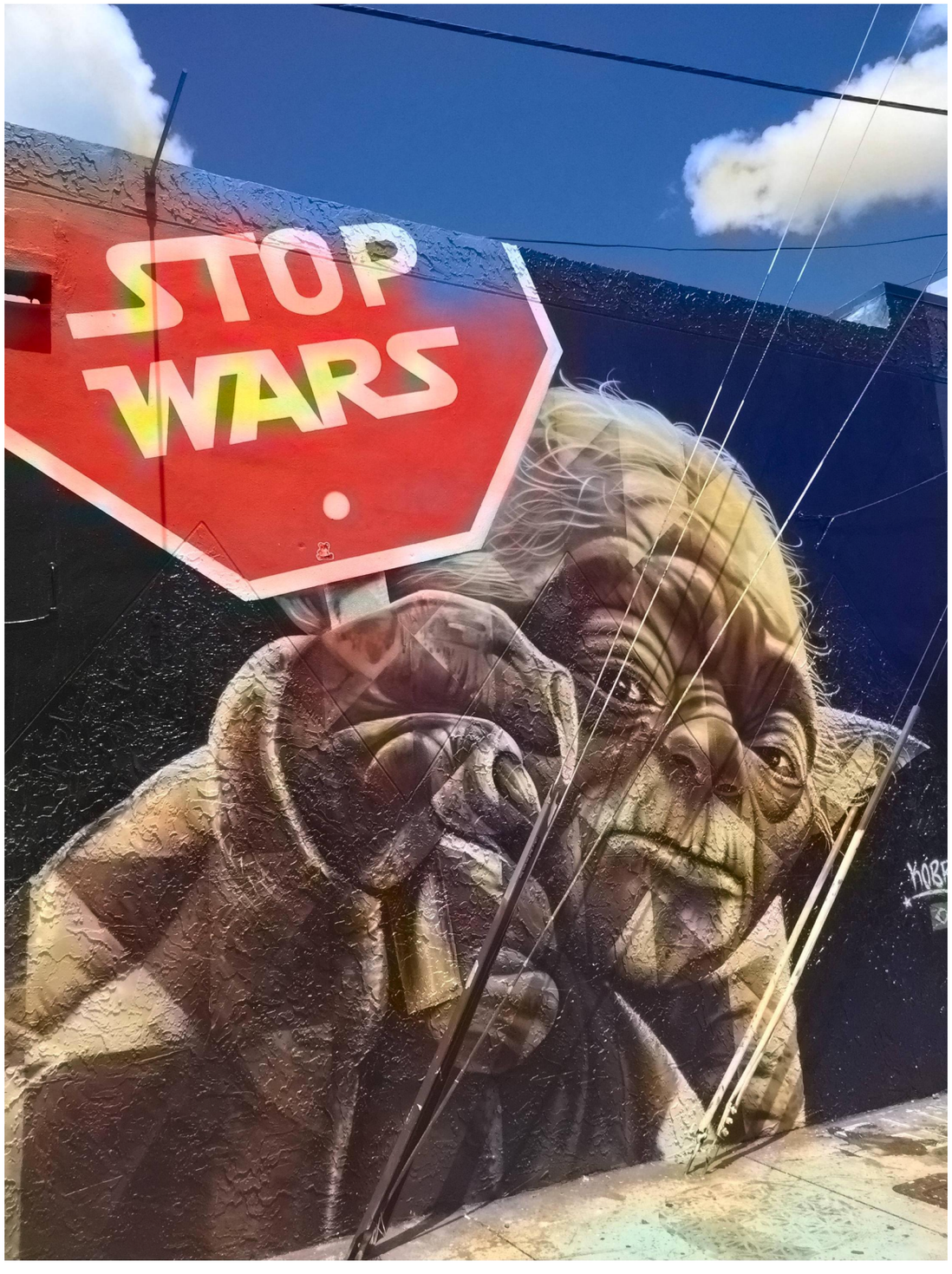
Biais



Reddit /u/SherySantucci

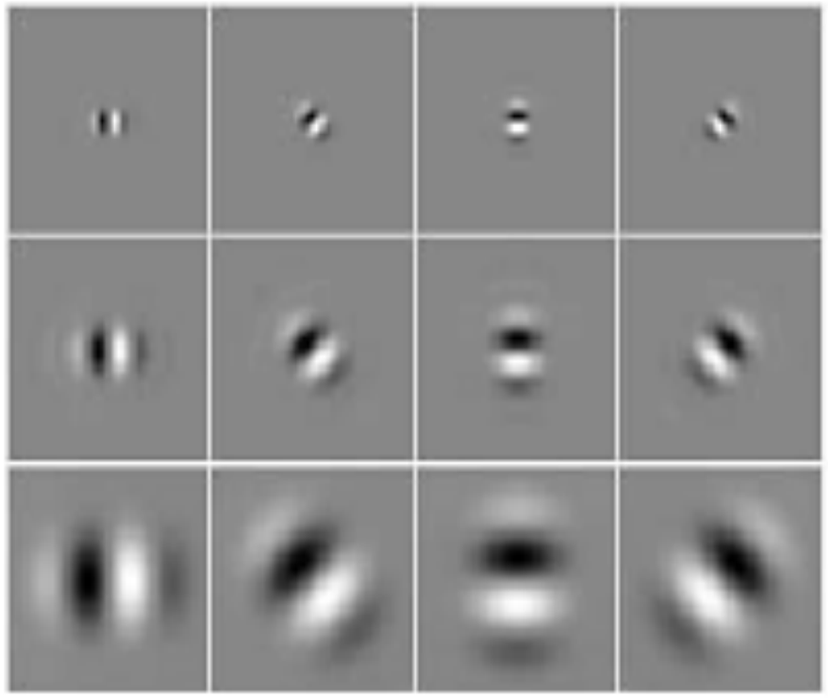
Reddit ColorizeBot

Photo de Reddit /u/Timteroo,
Murale de Eduardo Kobra



Reddit ColorizeBot

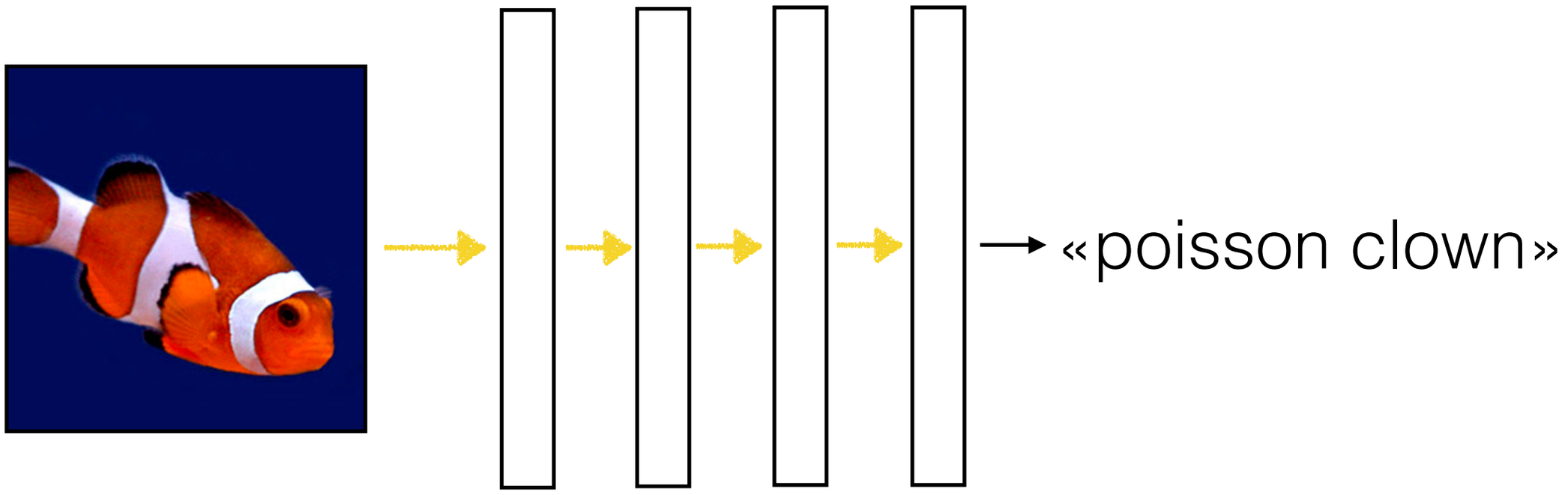
Représenter les textures



Pourquoi *ces* caractéristiques exactement?

A. Owens

Choix de la représentation

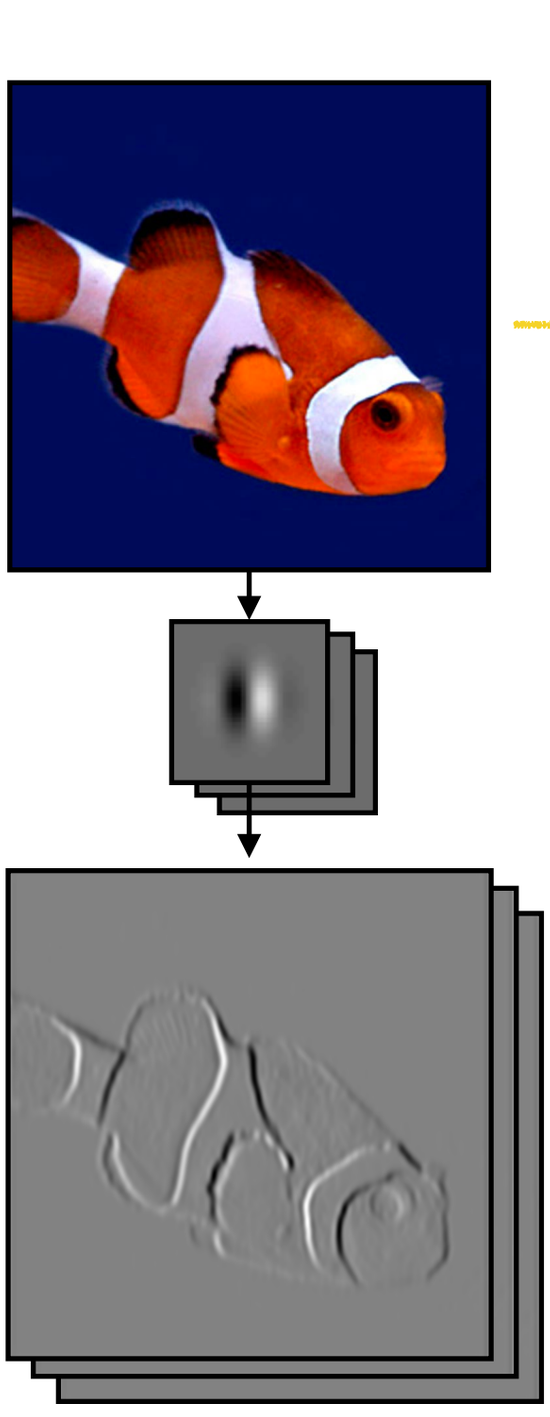


Au lieu de banques de filtres pré-déterminés, utilisons les caractéristiques apprises par un réseau de neurones!

Un réseau de neurone entraîné à reconnaître les objets devient un calculateur de caractéristiques très robustes.

A. Owens

Extraire les caractéristiques d'une image



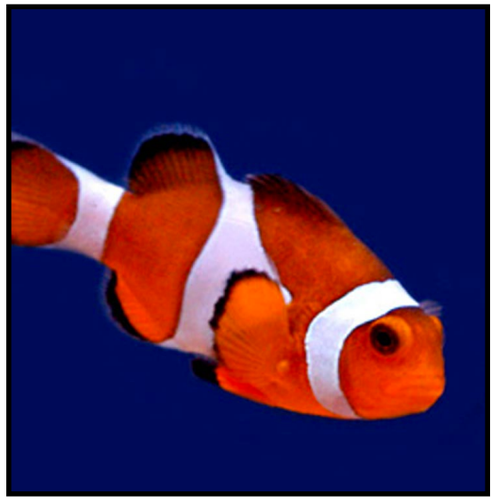
$I \cdot W_1$

convolution
(= multiplication matricielle)

A. Owens

Extraire les caractéristiques d'une image

«conv1»



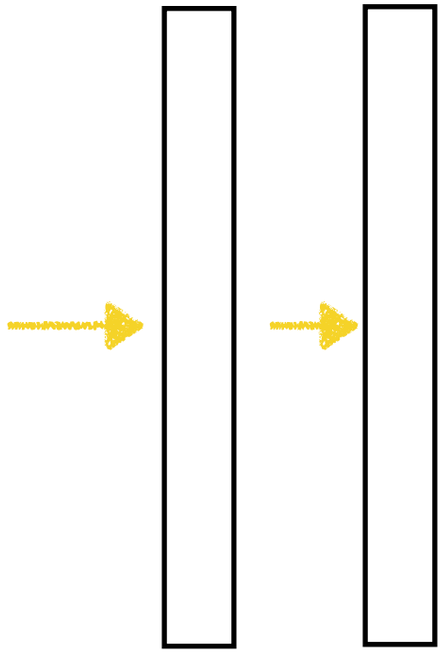
$\lfloor I \cdot W_1 \rfloor$ convolution suivie d'une non-linéarité

où $\lfloor \vec{x} \rfloor_i = \max(x_i, 0)$

i.e. valeurs négatives = 0

Extraire les caractéristiques d’une image

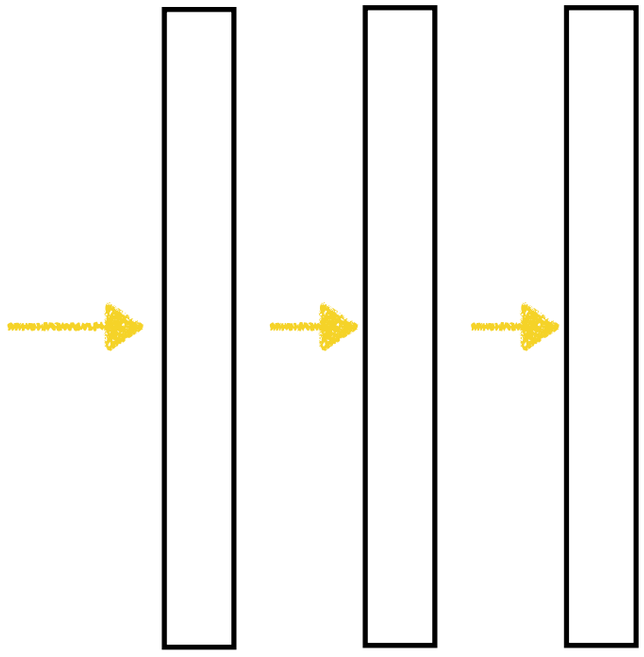
«conv2»



$$\lfloor \lfloor I \cdot W_1 \rfloor \cdot W_2 \rfloor$$

Extraire les caractéristiques d’une image

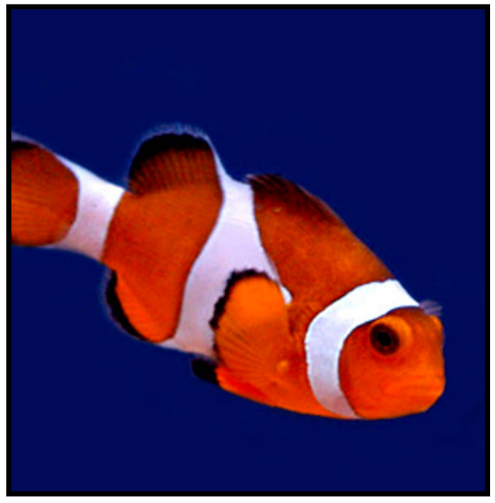
«conv3»

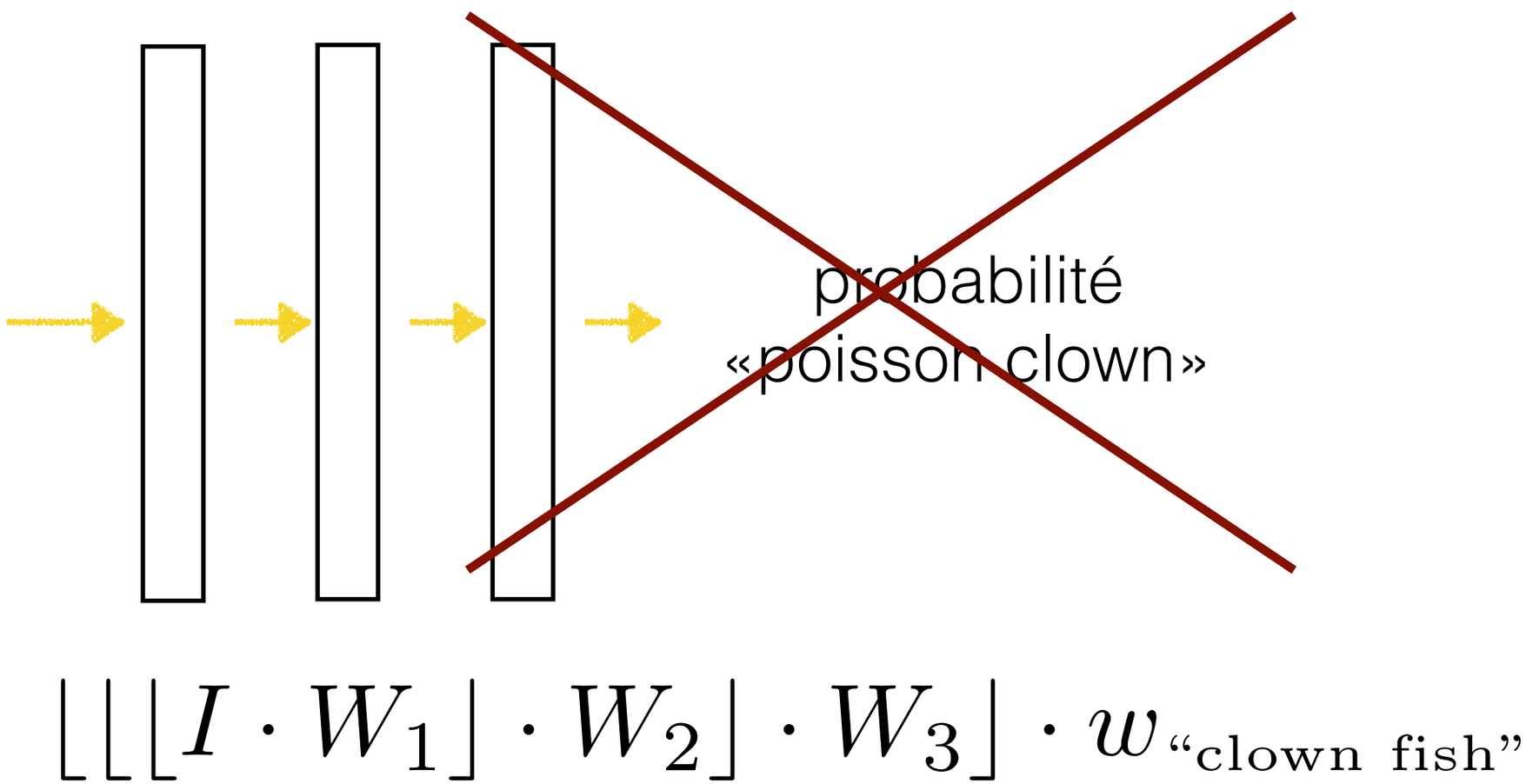


$$\lfloor \lfloor \lfloor I \cdot W_1 \rfloor \cdot W_2 \rfloor \cdot W_3 \rfloor$$

A. Owens

Extraire les caractéristiques d’une image

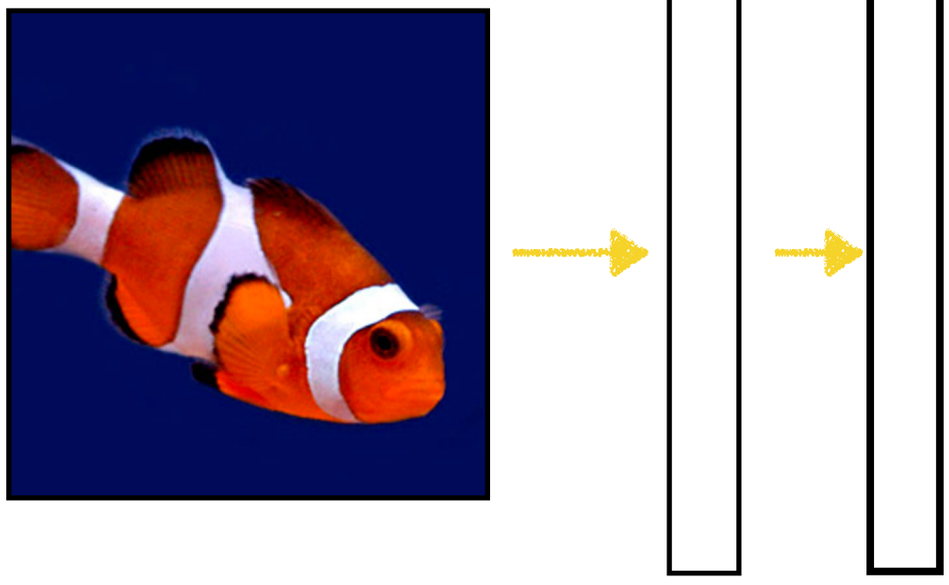




$$\lfloor \lfloor \lfloor I \cdot W_1 \rfloor \cdot W_2 \rfloor \cdot W_3 \rfloor \cdot w_{\text{“clown fish”}}$$

A. Owens

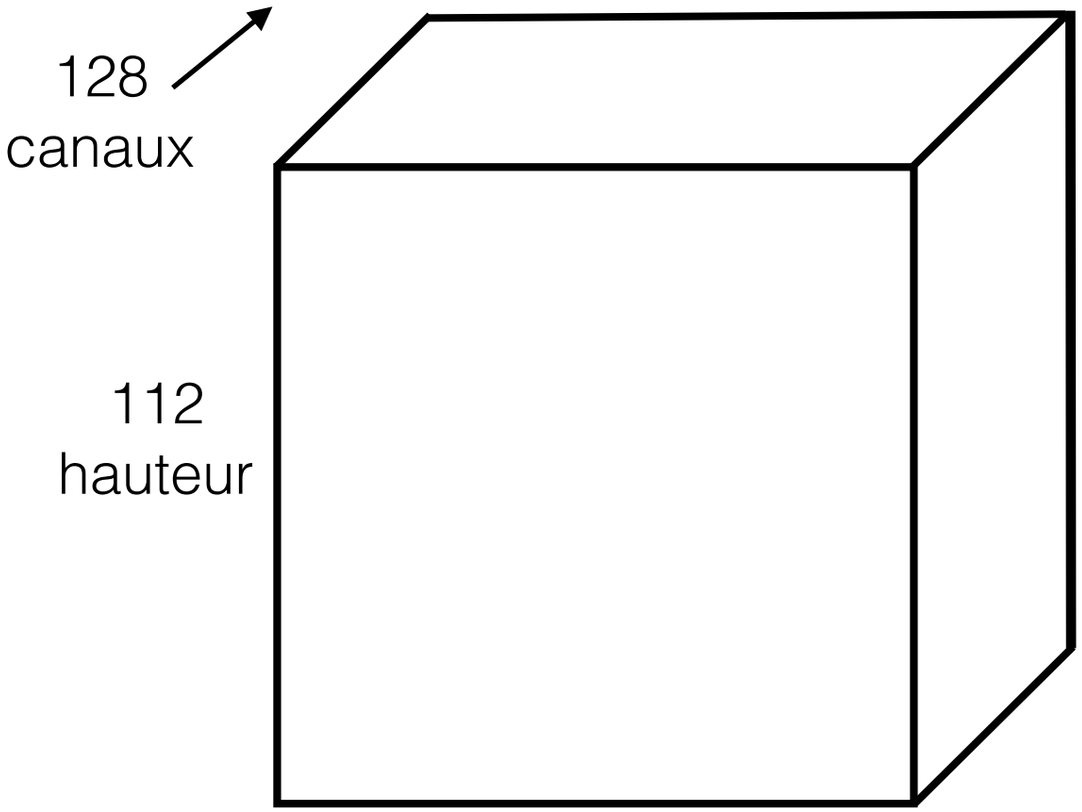
Extraire les caractéristiques d’une image



$$\lfloor \lfloor I \cdot W_1 \rfloor \cdot W_2 \rfloor$$

«activations» de la couche «conv2»

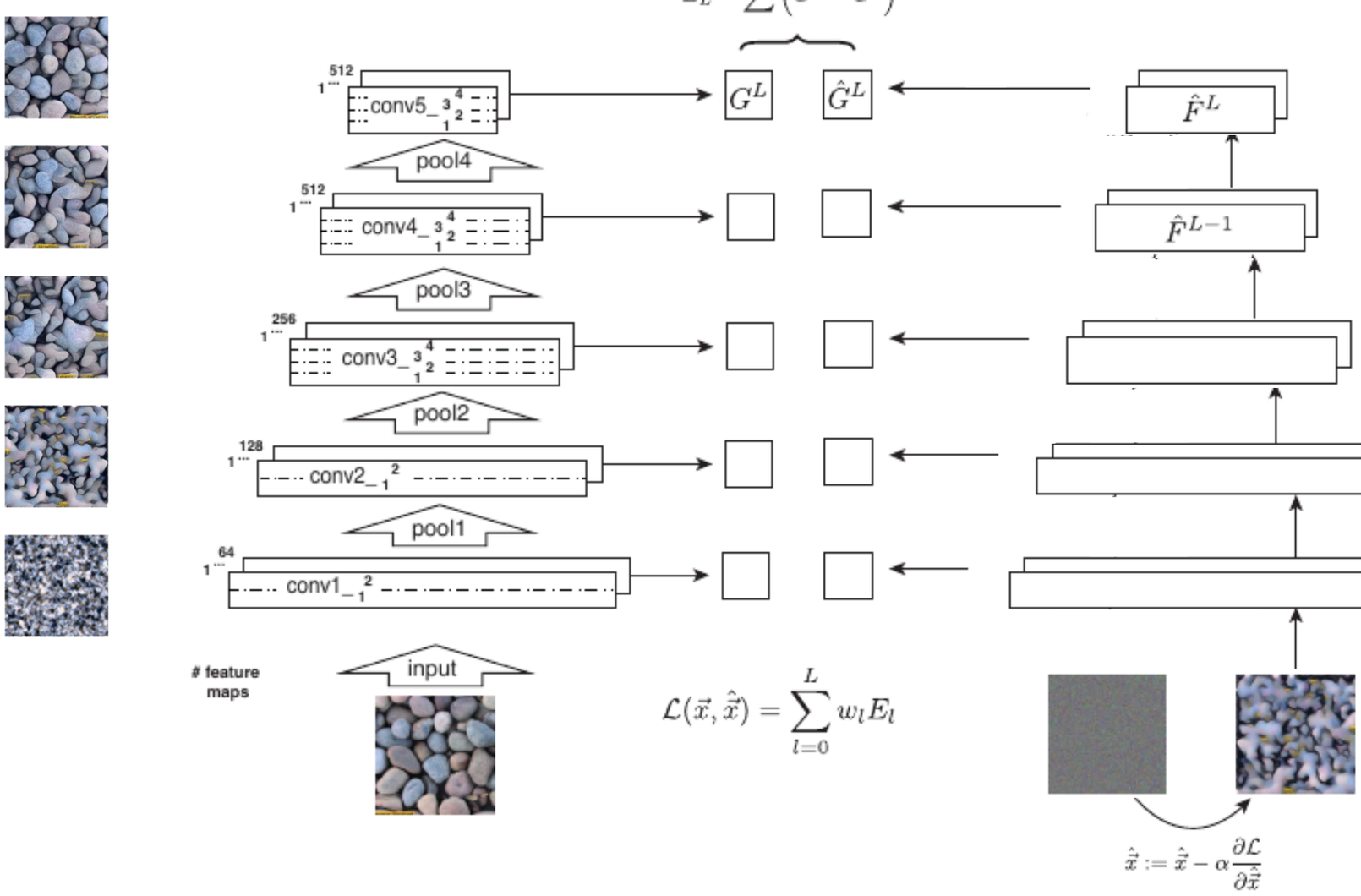
Extraire les caractéristiques d'une image



activations de «conv2»

e.g. (112 x 112) x 128 pour conv2 d'un modèle populaire (VGG19)

Synthèse de textures par CNN



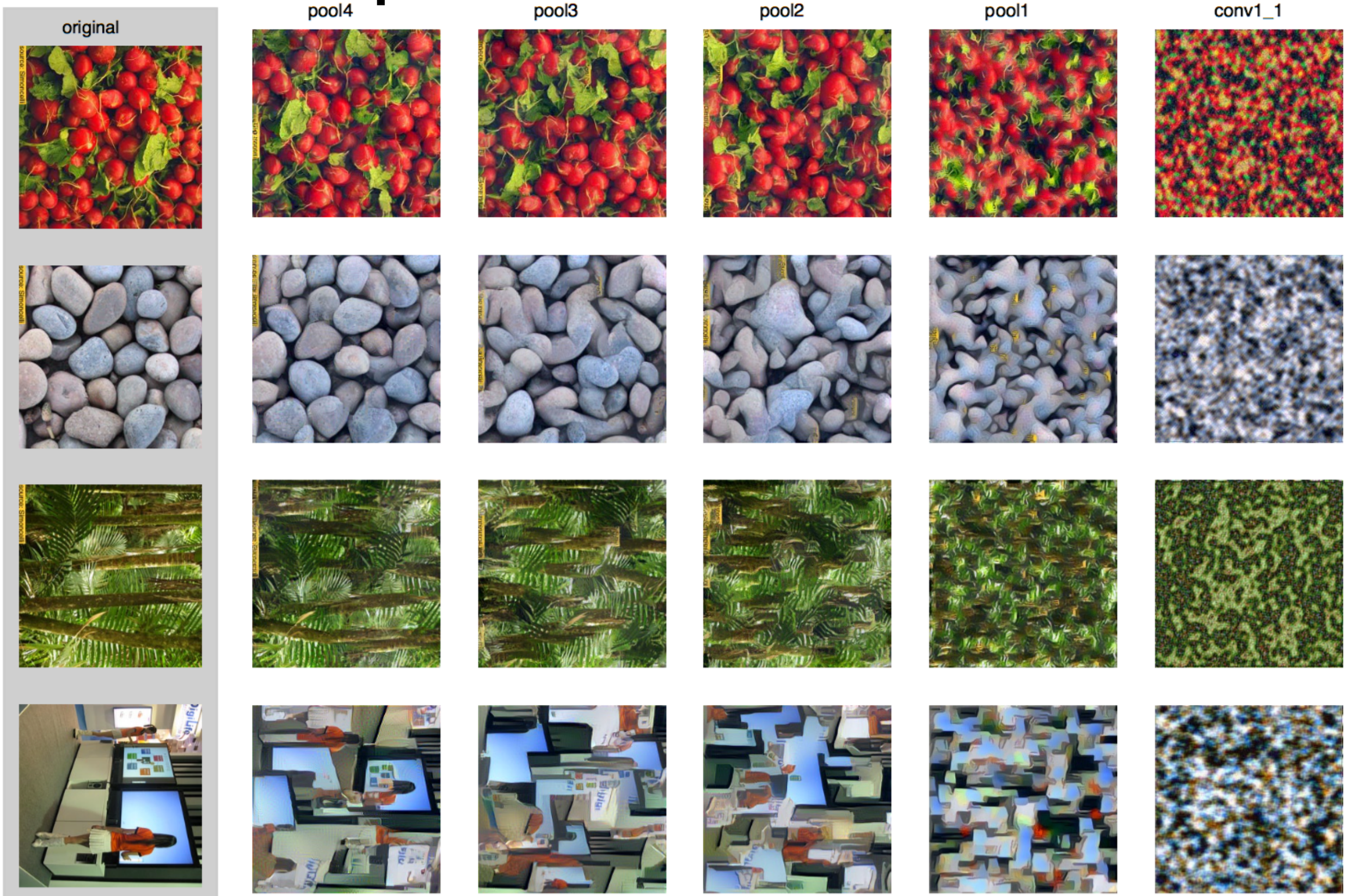
J. Zhang

Synthèse de textures par CNN

[Gatys et al. 2005]

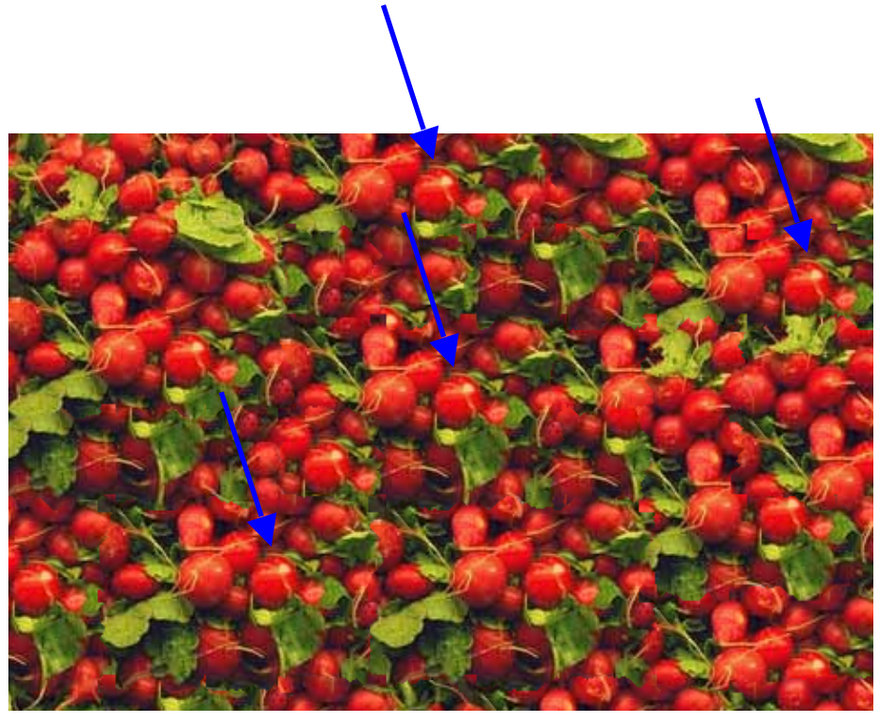
Haute → Basse

Statistiques de textures



Synthèse de textures par CNN

Échantillonner la texture originale



[Gatys et al. 2005]

Synthèse de textures par CNN

Synthesis

Source

[Gatys et al. 2005]

Synthèse de textures par CNN

Synthesis

Source

[Gatys et al. 2005]

Capturer le style artistique

Comment transférer le style d'une peinture vers une photo?

[Gatys et al. 2016]

Capturer le style artistique

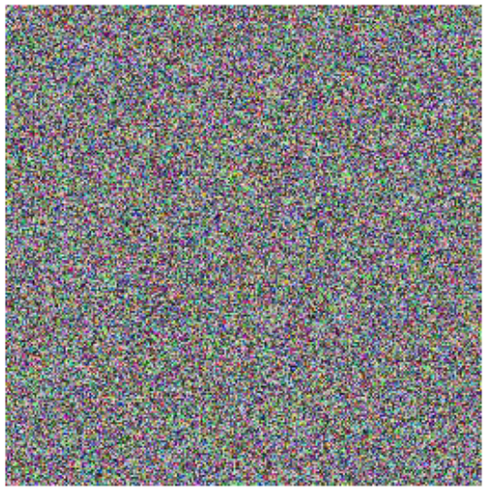
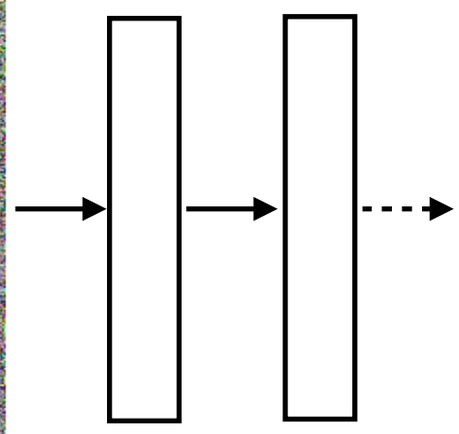


Image de synthèse



$\hat{G}$

$$\sum_{i,j}(\hat{G}_{ij} - G_{ij})^2$$

Rajouter le contenu de la photo

$$\sum_{i}\sum_{x,y}(c_i(x,y) - \hat{c}_i(x,y))^2$$

$c_i(x,y)$

$\hat{c}_i(x,y)$

A. Owens

A. Owens

A. Owens

A. Owens

A. Owens

Londres le jour

New York le soir

Londres le soir?