UNIVERSITÉ LAVAL

GLO-4030/7030 APPRENTISSAGE PAR RÉSEAUX DE NEURONES PROFONDS

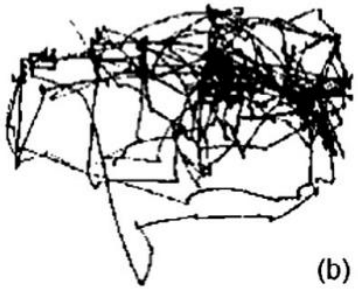
Attention

(image et texte)

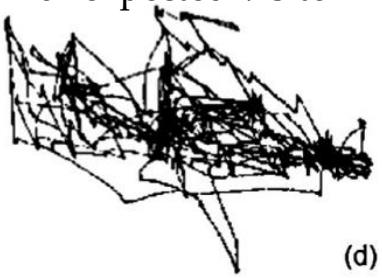
Philippe Giguère

Attention visuelle humaine

Position du regard en fonction de la question posée

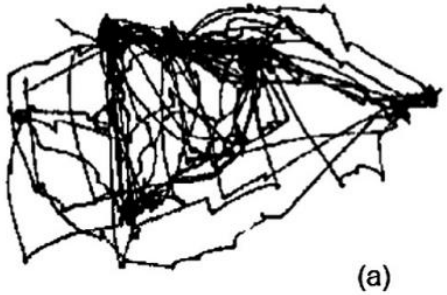


Estimate the wealth of the family



Summarize what the family had been doing before the arrival of the ''unexpected visitor''

Remember the position of the people and objects in the room



No specific task

Give the ages of the people

Remember the clothes worn by the people

Estimate how long the ''unexpected visitor'' had been away from the family

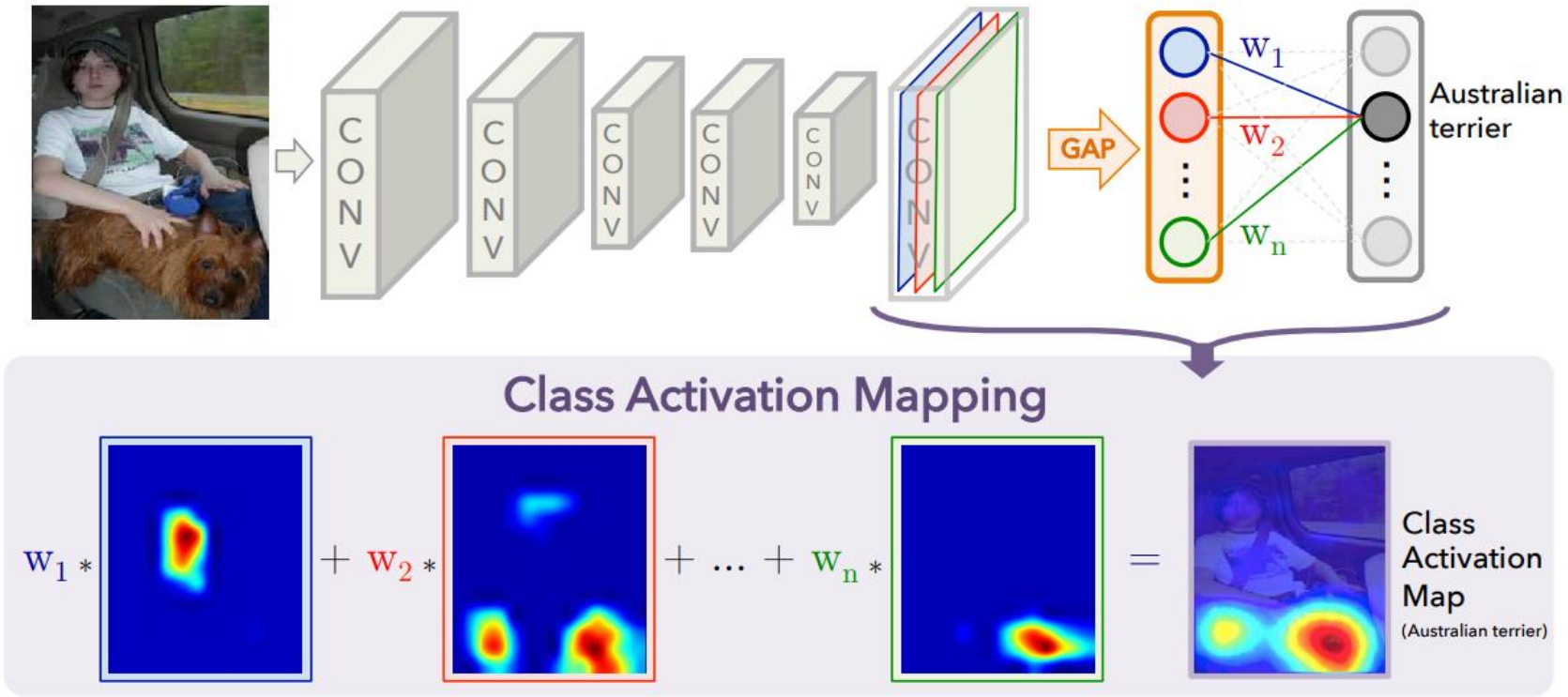
Yarbus, A. (1967). Eye movements and vision. New York: Plenum Press (Translated from the Russian edition by Haigh, B).

Attention visuelle humaine

- Fovéa dans l'oeil

Globlal average pooling

- Vers la localisation et l'attention visuelle

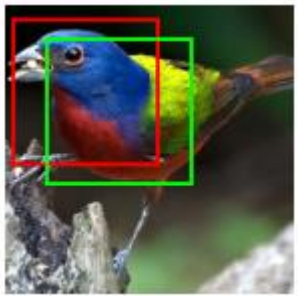


- Donne une certaine interprétabilité aux résultats

Zhou et al., Learning Deep Features for Discriminative Localization, CVPR 2016.

Spatial transformer : attention

2 parallel ST



Spatial Transformer

Image captioning

- Attention séquentielle sur l'image

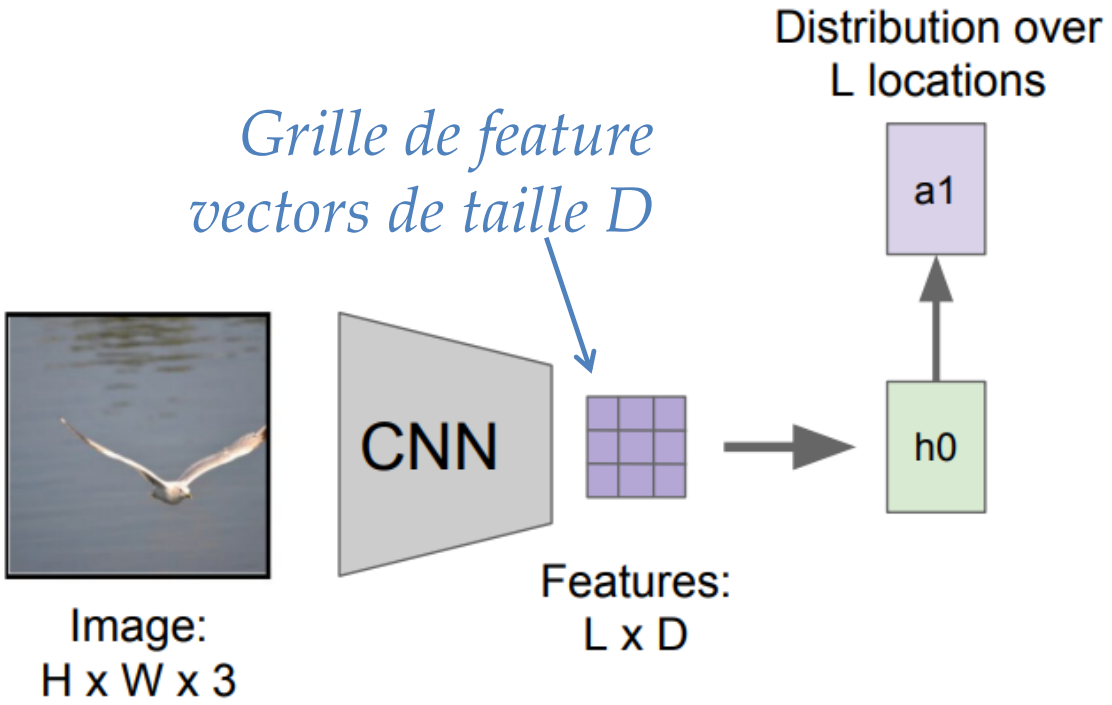
A dog is running in the grass with a frisbee

A woman is holding a cat in her hand

A cat is sitting on a tree branch

A man in a baseball uniform throwing a ball

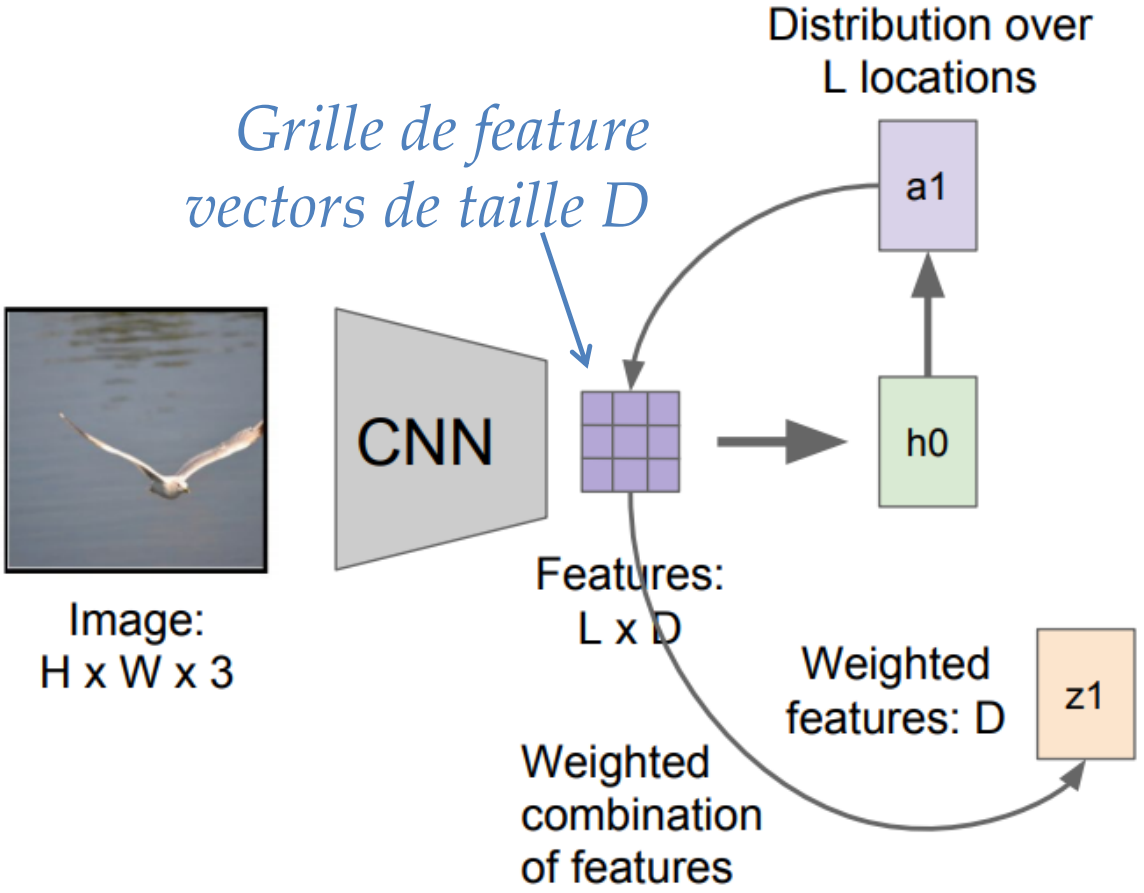
Image captioning avec attention



Xu et al., Show, Attend and Tell:
Neural Image Caption Generation
with Visual Attention, ICML 2015.

Tiré de cs231n

Image captioning avec attention



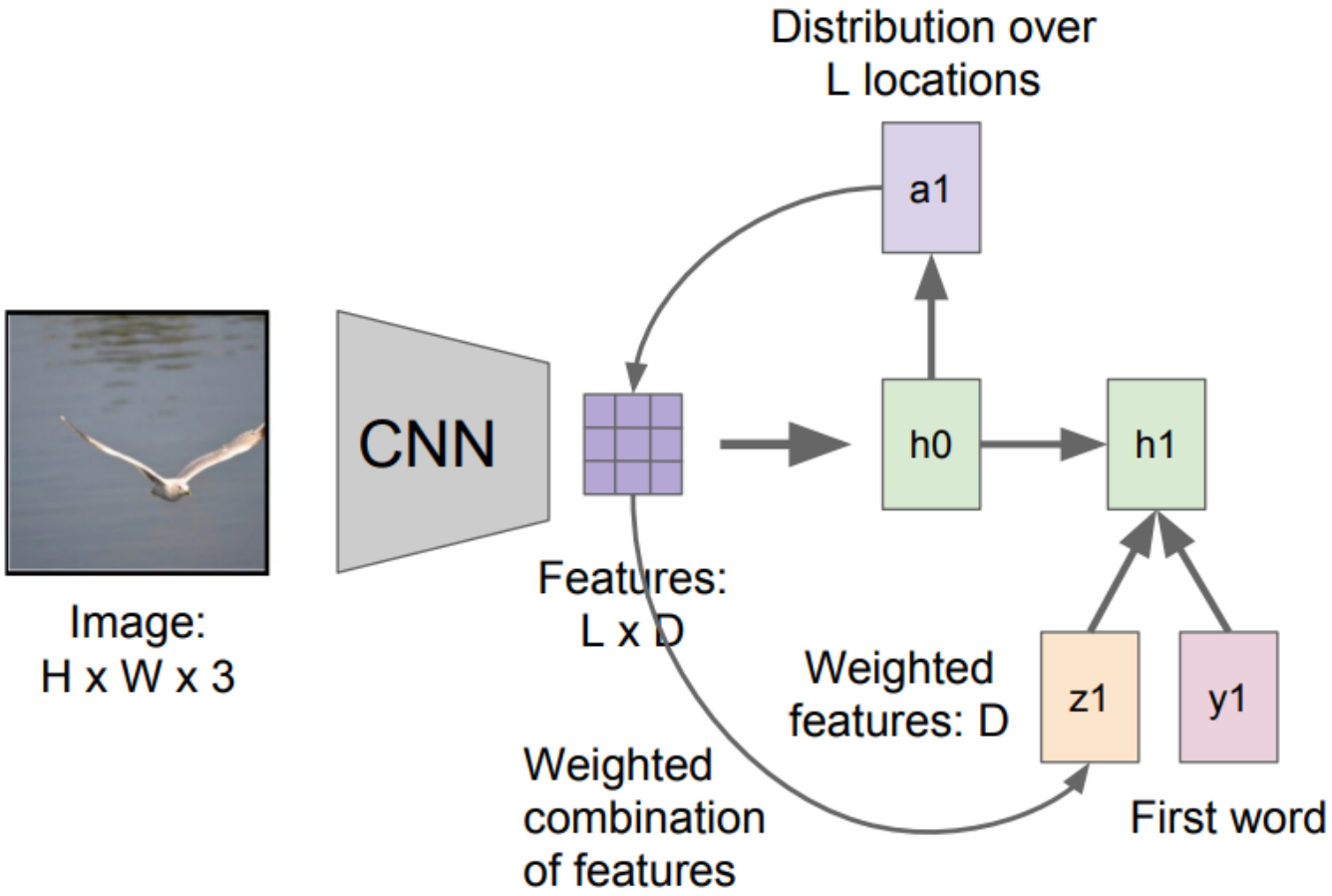
$$z = \sum_{i=1}^{L} p_i v_i$$

(soft attention)

Xu et al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML 2015.

Tiré de cs231n

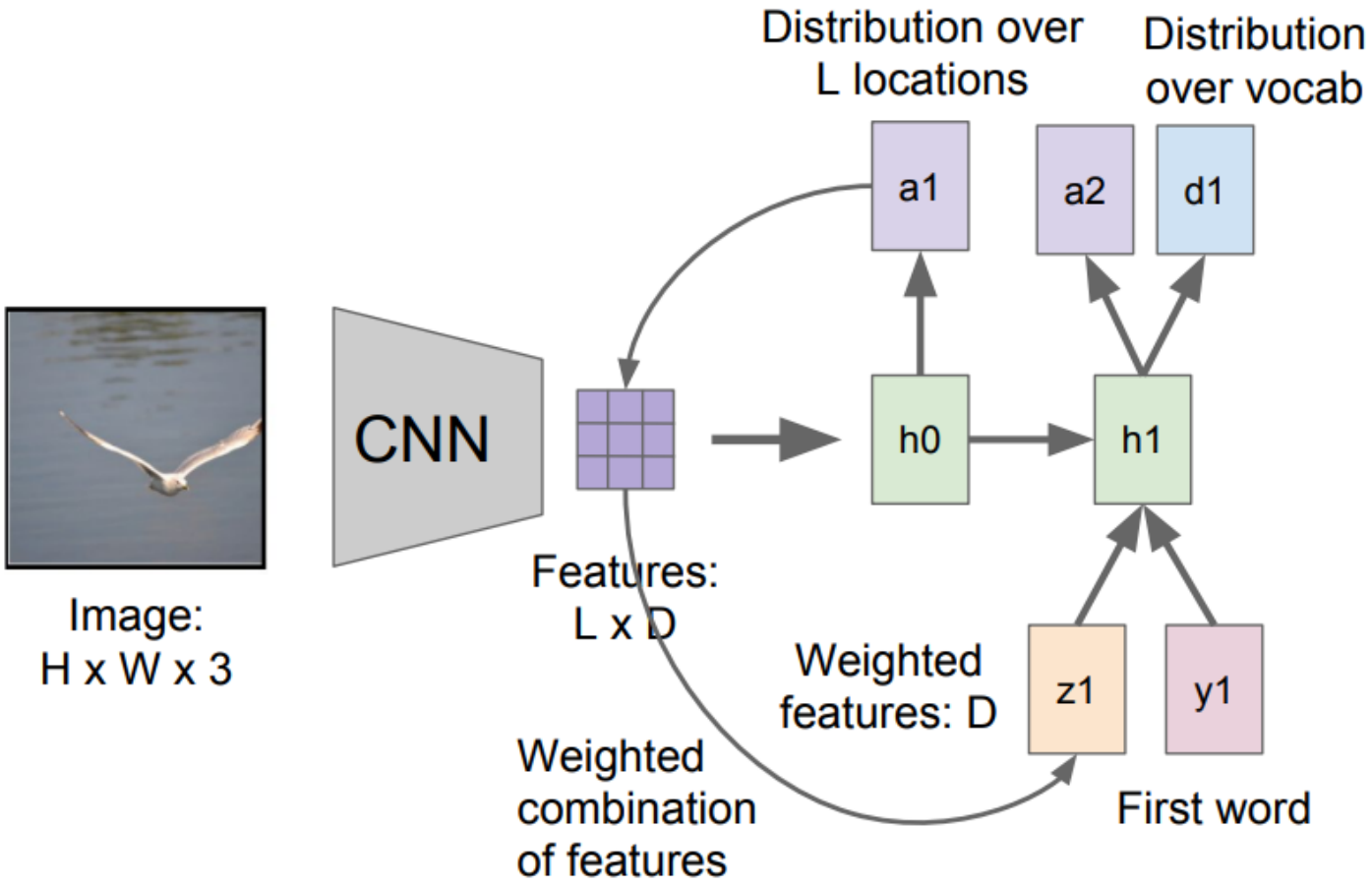
Image captioning avec attention



Xu et al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML 2015.

Tiré de cs231n

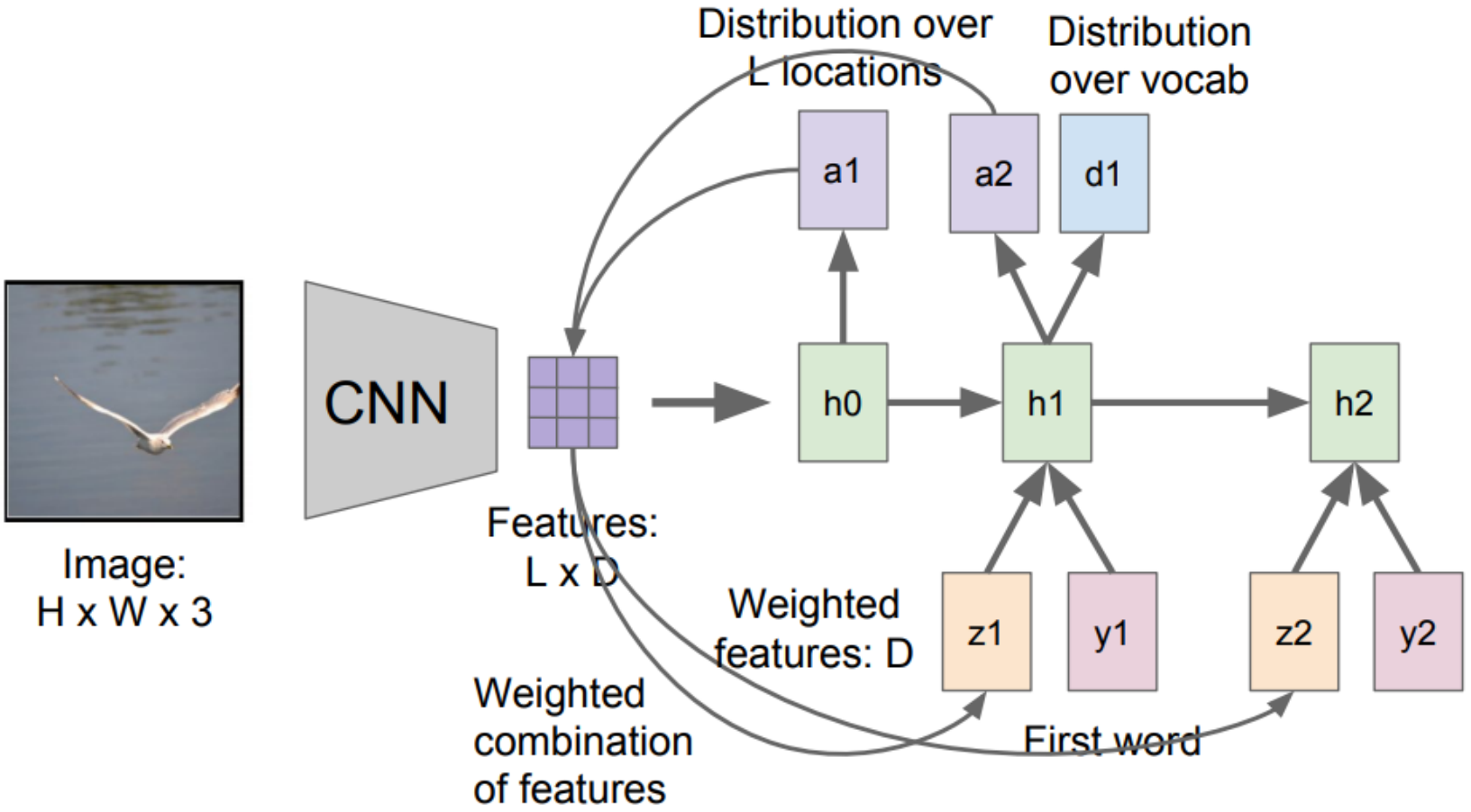
Image captioning avec attention



Xu et al., Show, Attend and Tell:
Neural Image Caption Generation
with Visual Attention, ICML 2015.

Tiré de cs231n

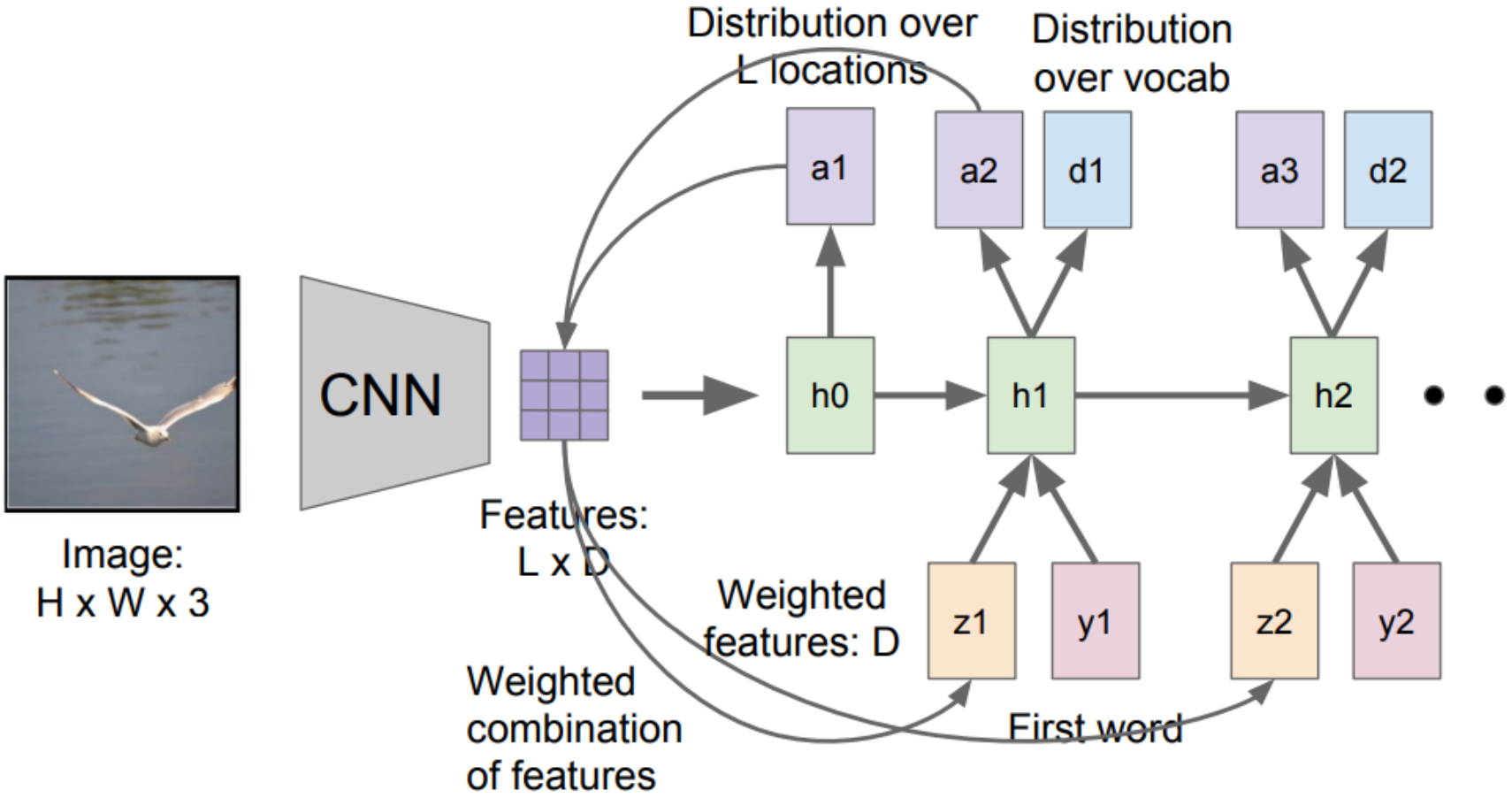
Image captioning avec attention



Xu et al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML 2015.

Tiré de cs231n

Image captioning avec attention



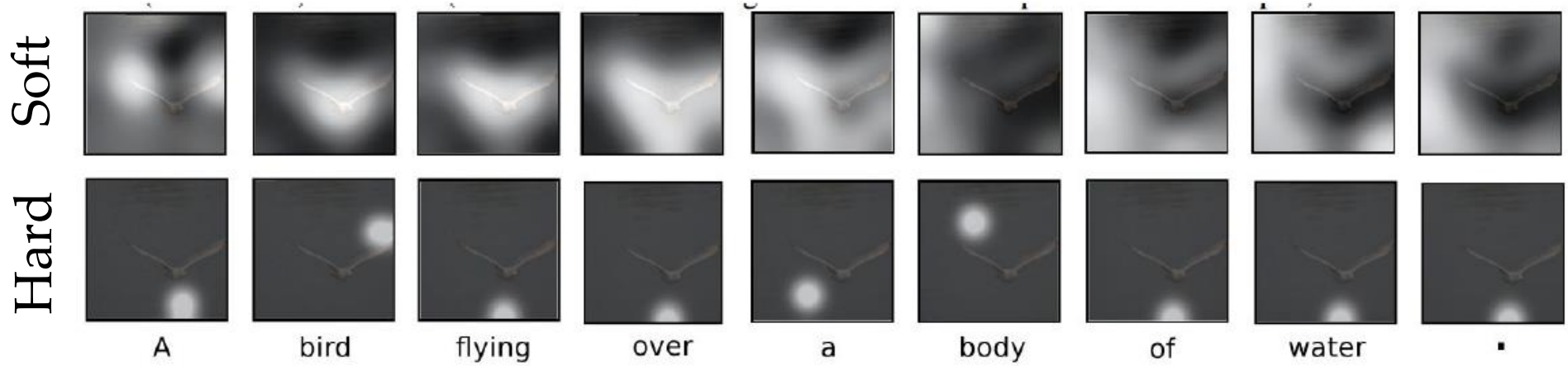
Xu et al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML 2015.

Tiré de cs231n

Soft vs. hard attention

- Soft
  - Sommes pondérées
  - Poids calculés par une softmax (cas d'utilisation qui n'est pas en sortie)
  - dérivable end-to-end
- Hard
  - Softmax : distribution de probabilité de piger
  - pige un élément sur lequel diriger l'attention
  - non-dérivable + difficile à entrainer (question de la semaine passée sur VAE)

Soft vs. hard attention

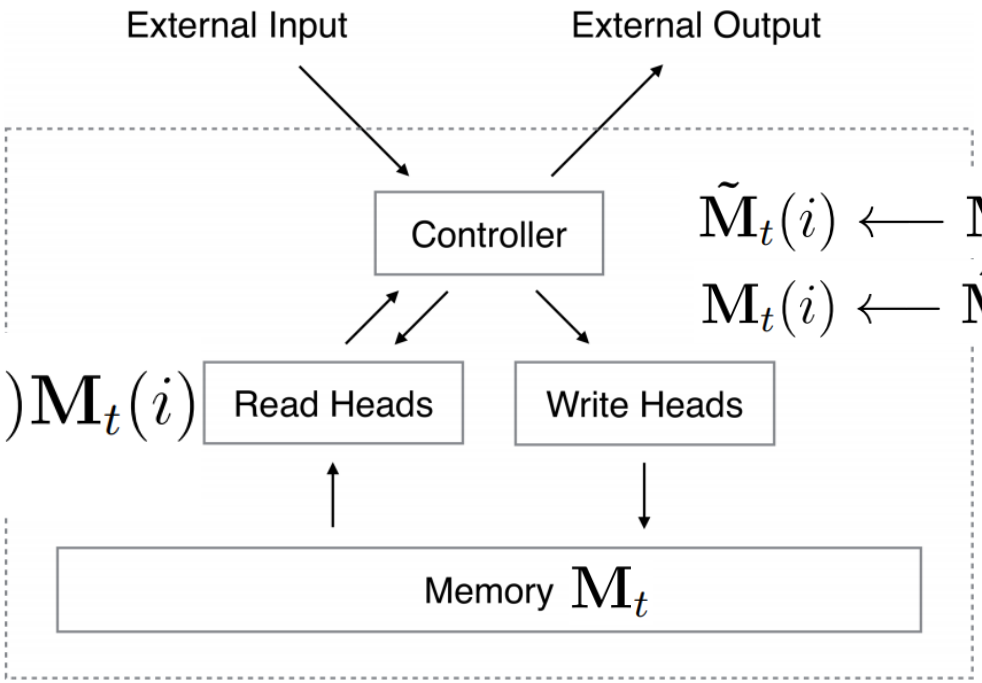


Neural Turing Machines (Oct. 2014)

Alex Graves gravesa@google.com
Greg Wayne gregwayne@google.com
Ivo Danihelka danihelka@google.com

- "Ordinateur" dérivable end-to-end
- Séparation calcul/mémoire
- Head : attention

$$w_t^c(i) \longleftarrow \frac{\exp\left(\beta_t K\left[\mathbf{k}_t, \mathbf{M}_t(i)\right]\right)}{\sum_j \exp\left(\beta_t K\left[\mathbf{k}_t, \mathbf{M}_t(j)\right]\right)}$$



$$\mathbf{r}_t \longleftarrow \sum_i w_t(i)\mathbf{M}_t(i)$$

$$\tilde{\mathbf{M}}_t(i) \longleftarrow \mathbf{M}_{t-1}(i)\left[\mathbf{1} - w_t(i)\mathbf{e}_t\right]$$

$$\mathbf{M}_t(i) \longleftarrow \tilde{\mathbf{M}}_t(i) + w_t(i)\,\mathbf{a}_t$$

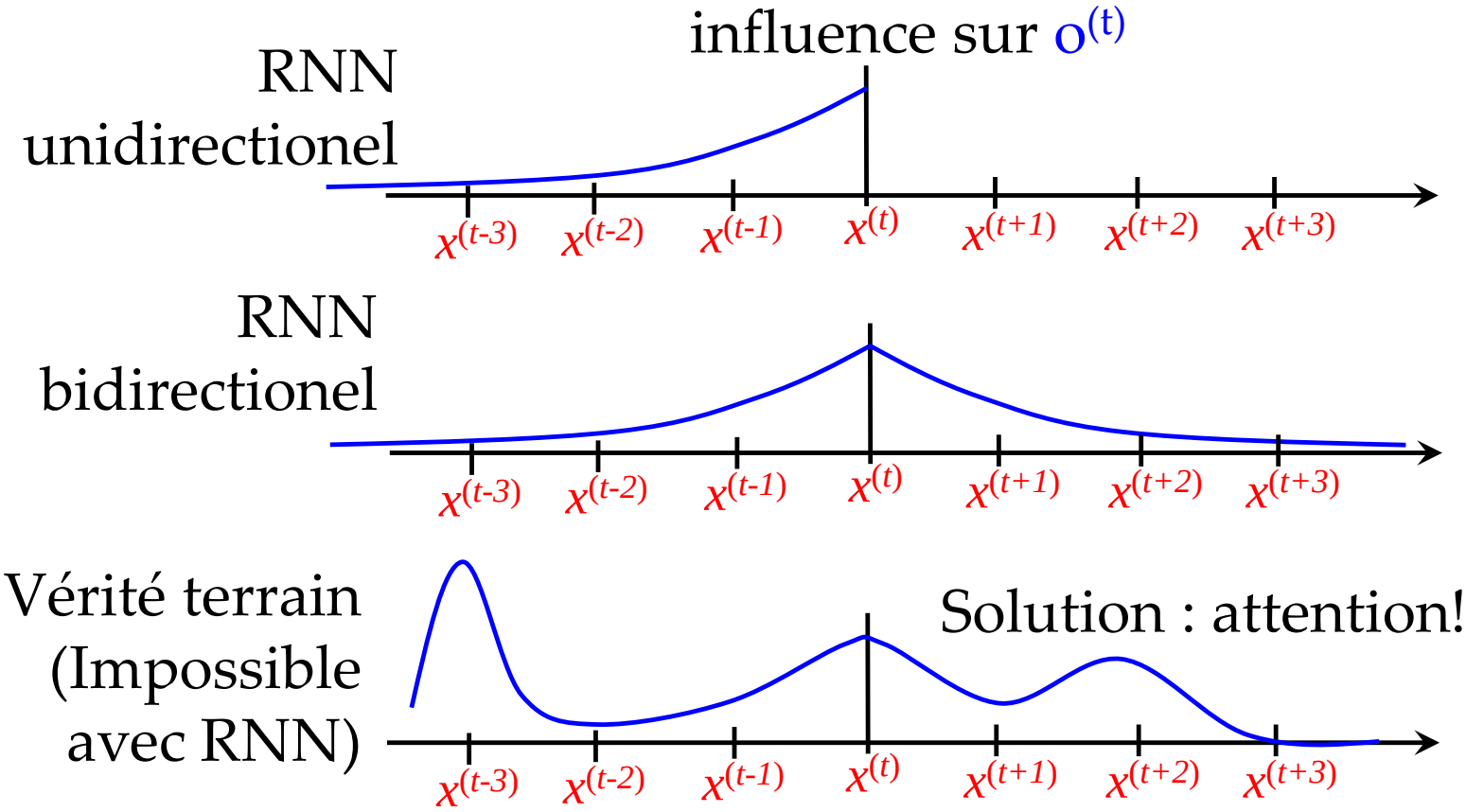
vecteur d'effacement

vecteur d'addtion

(LSTM)

Rappel : longue portée

- Influence à longue portée difficile dans RNN
- RNN : décroissance exponentielle de l'influence



Published as a conference paper at ICLR 2015

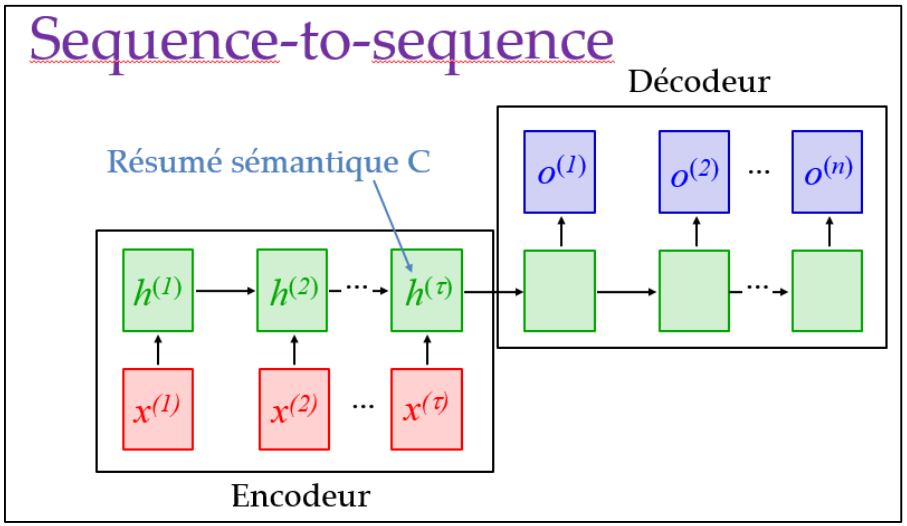
NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

Dzmitry Bahdanau
Jacobs University Bremen, Germany

KyungHyun Cho **Yoshua Bengio***
Université de Montréal

Attention pour traduction

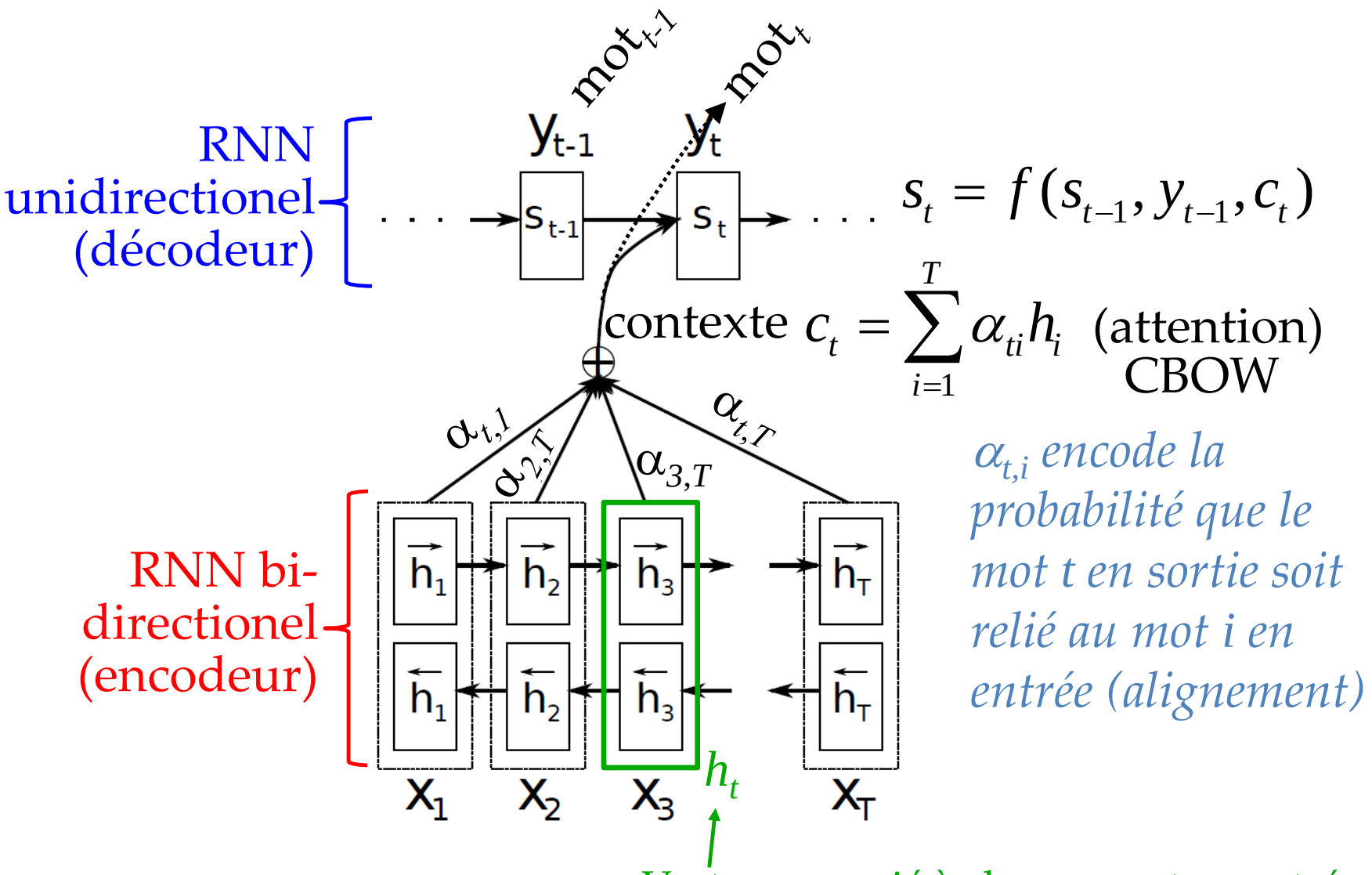
- Résumé sémantique d'une phrase en un seul vecteur <u>est trop restrictif</u>



- Propose plutôt d'associer un vecteur supplémentaire (état caché) à chaque mot
- Mécanisme d'**attention** *soft* sur les états des mots en entrée pour aider à la prédiction en sortie
- Généralise mieux pour des phrases longues

Bahdanau et al., Neural Machine Translation by Jointly Learning to Align and Translate, ICLR 2015.

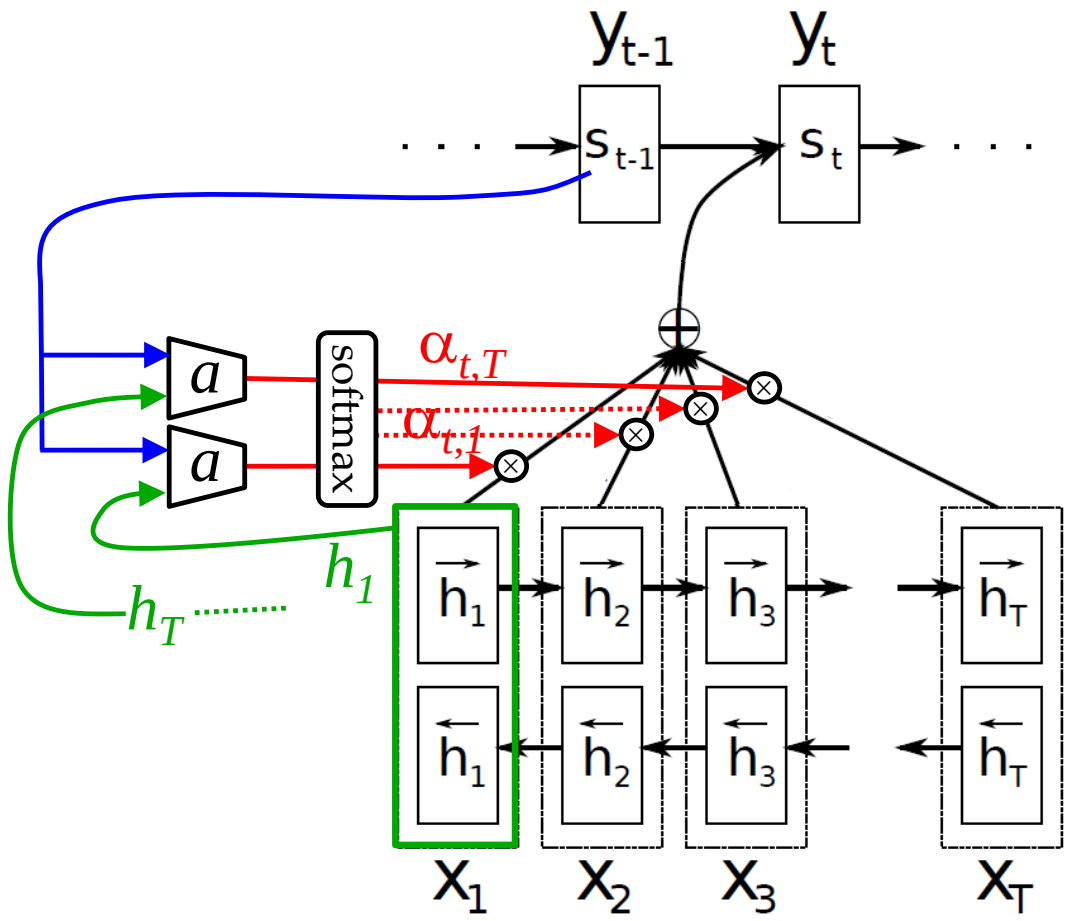
Architecture



$\alpha_{t,i}$ encode la probabilité que le mot t en sortie soit relié au mot i en entrée (alignement)

Architecture : réseau *a* d'attention

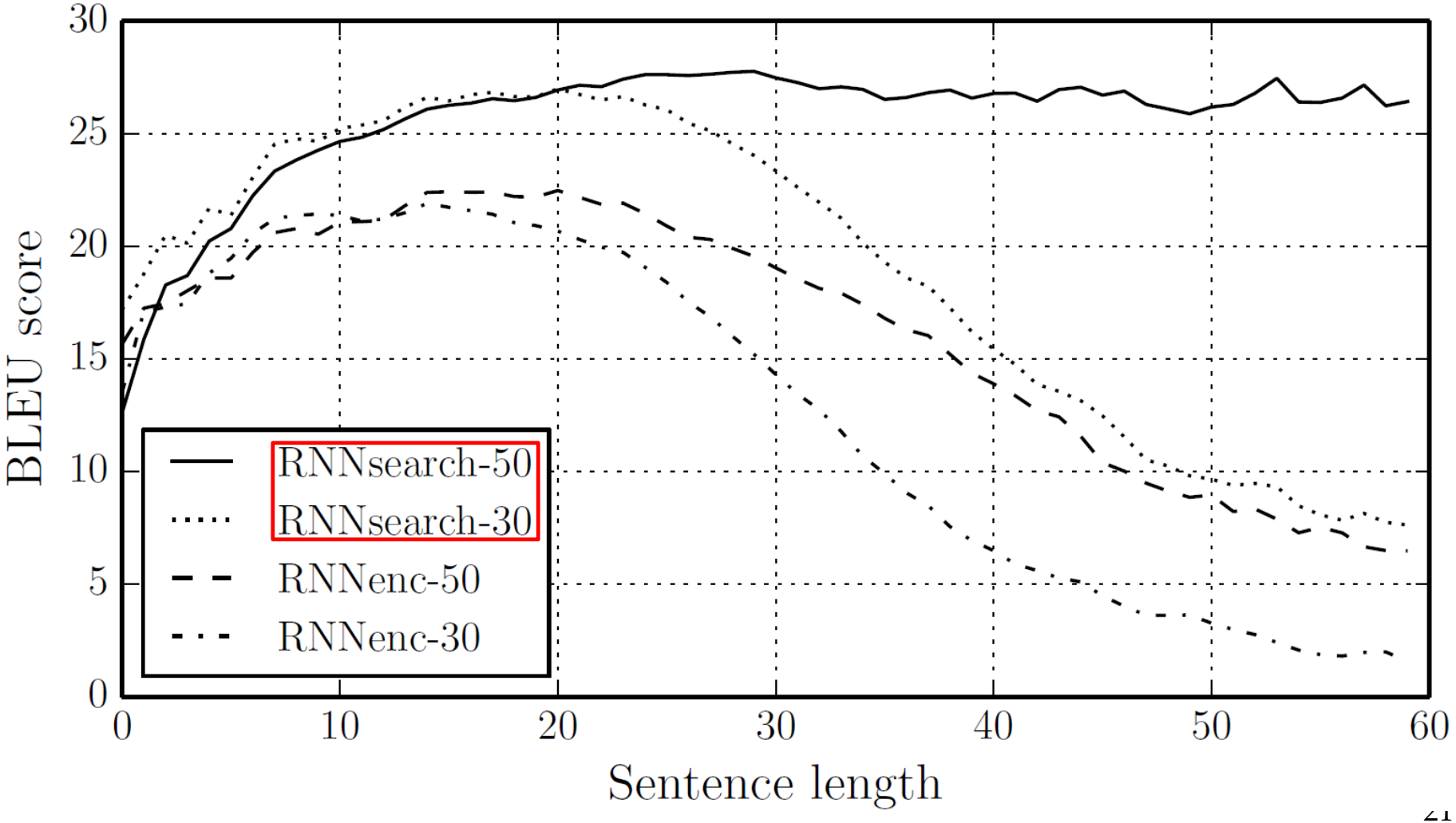
Réseau *a* peu profond



Bahdanau et al., Neural Machine Translation by Jointly Learning to Align and Translate, ICLR 2015.

Résultats

- Fonctionne bien pour de longues phrases



Exemple alignement

- Pour le choix de l'article {le, la, l'}, le réseau regarde un mot en avant

Donne une certaine interprétabilité aux résultats

- Inversion de l'ordre des mots pour l'adjectif