

Réseaux de neurones convolutionnels (CNN) et apprentissage profond par renforcement (DRL)

Arthur Aubret

Université Lyon 1

10/12/2020 et 17/12/2020

Sommaire

1 Réseaux de neurones convolutionnels

- Introduction
- Masque convolution 1D
- Masque convolution 2D
- Autres types de données: brève introduction

2 Apprentissage profond par renforcement

- Rappels
- DQN
- Policy-based methods
- Actor-Critic

3 Motivation intrinsèque

- Problèmes du RL
- Récompenses intrinsèque
- Différents types de récompense intrinsèque

Rappels: perceptron multi-couches

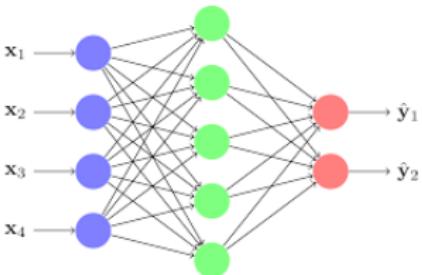


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- Couche entièrement connectée.

Rappels: perceptron multi-couches

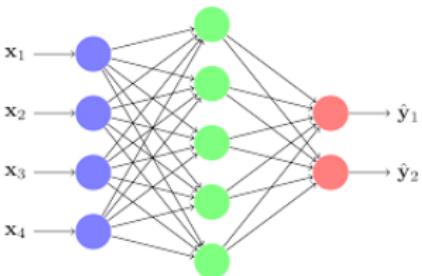


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- Couche entièrement connectée.
- Activation non linéaire.

Rappels: perceptron multi-couches

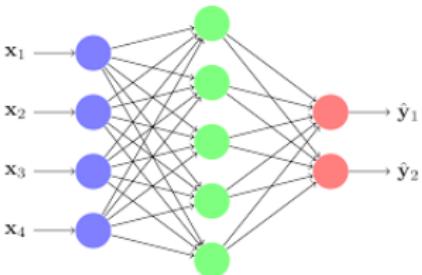


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- Couche entièrement connectée.
- Activation non linéaire.
- Et on répète...

Introduction

Rappels: apprentissage

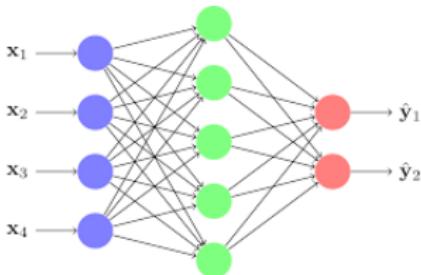


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- ### ❶ Génération aléatoire des poids W du modèle F_W .

Introduction

Rappels: apprentissage

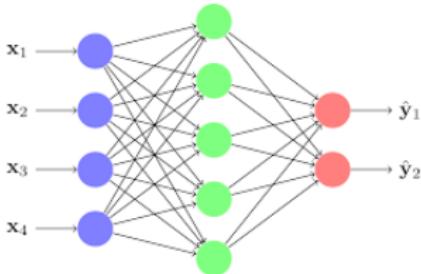


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- ➊ Génération aléatoire des poids W du modèle F_W .
 - ➋ On passe nos données dans le modèle: $\hat{Y}_{train} = F_W(X_{train})$.

Introduction

Rappels: apprentissage

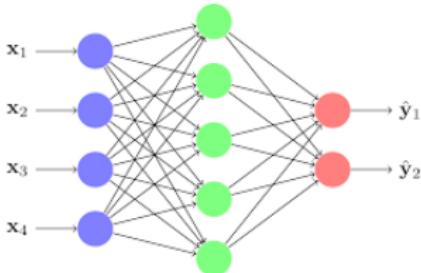


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- ➊ Génération aléatoire des poids W du modèle F_W .
 - ➋ On passe nos données dans le modèle: $\hat{Y}_{train} = F_W(X_{train})$.
 - ➌ On calcule la fonction de coût $L(\hat{Y}_{train}, Y_{train})$.

Introduction

Rappels: apprentissage

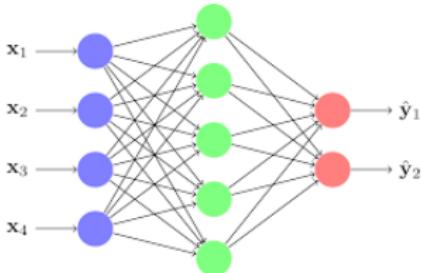


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- ➊ Génération aléatoire des poids W du modèle F_W .
 - ➋ On passe nos données dans le modèle: $\hat{Y}_{train} = F_W(X_{train})$.
 - ➌ On calcule la fonction de coût $L(\hat{Y}_{train}, Y_{train})$.
 - ➍ On applique la backpropagation du gradient: $\nabla W = \frac{\partial L}{\partial W}$.

Introduction

Rappels: apprentissage

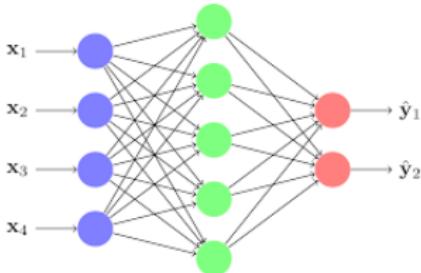


Figure: Réseau de neurone entièrement connecté.

<https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814616-explorez-les-reseaux-de-neurones-en-couches>

- ➊ Génération aléatoire des poids W du modèle F_W .
 - ➋ On passe nos données dans le modèle: $\hat{Y}_{train} = F_W(X_{train})$.
 - ➌ On calcule la fonction de coût $L(\hat{Y}_{train}, Y_{train})$.
 - ➍ On applique la backpropagation du gradient: $\nabla W = \frac{\partial L}{\partial W}$.
 - ➎ On met à jour les poids: $W = W - \alpha \nabla W$.

Introduction

Explosion du nombre de paramètres

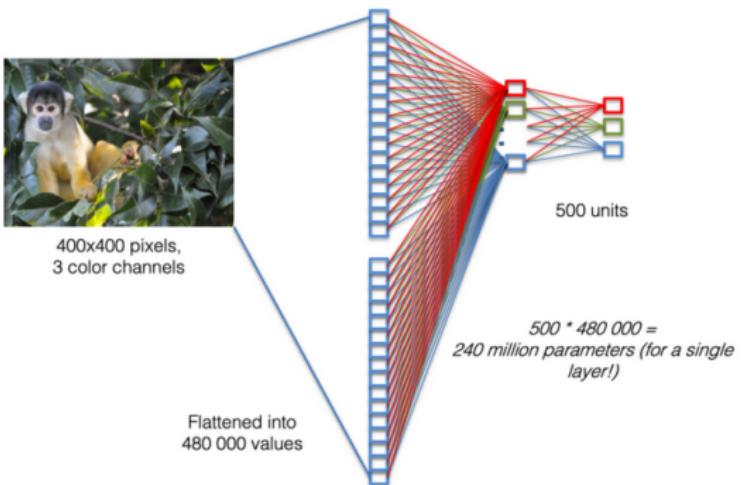


Figure: Surplus de paramètres lorsqu'un MLP est utilisé sur une image.

<https://chriswolfvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Variance par translation



On doit réapprendre plusieurs fois la même chose !

<https://chriswolfvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Réseau convolutionnel

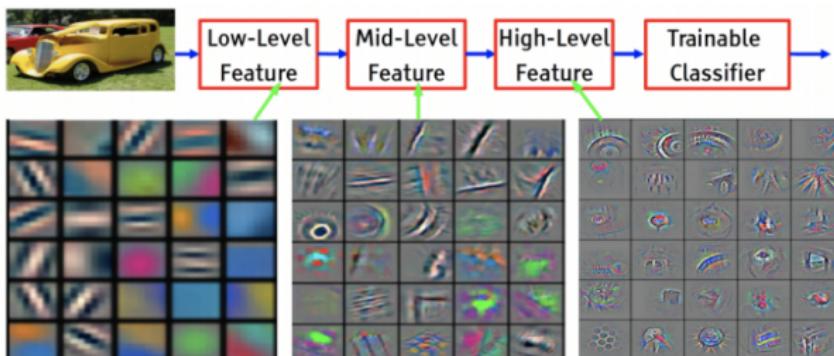


Figure: Réseau de neurones convolutionnel 2D

- Extraction de patterns locaux avec des filtres.

Réseau convolutionnel

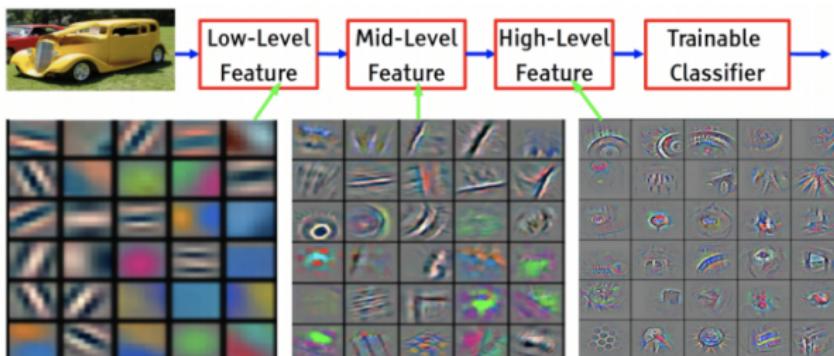


Figure: Réseau de neurones convolutionnel 2D

- Extraction de patterns locaux avec des filtres.
- Découverte automatique des patterns à chercher.

Introduction

Réseau convolutionnel

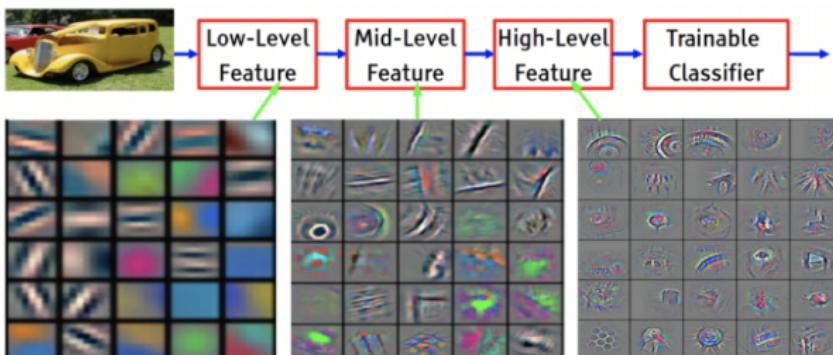


Figure: Réseau de neurones convolutionnel 2D

- Extraction de patterns locaux avec des filtres.
- Découverte automatique des patterns à chercher.
- ① Découverte des orientations des traits et couleurs.

Introduction

Réseau convolutionnel

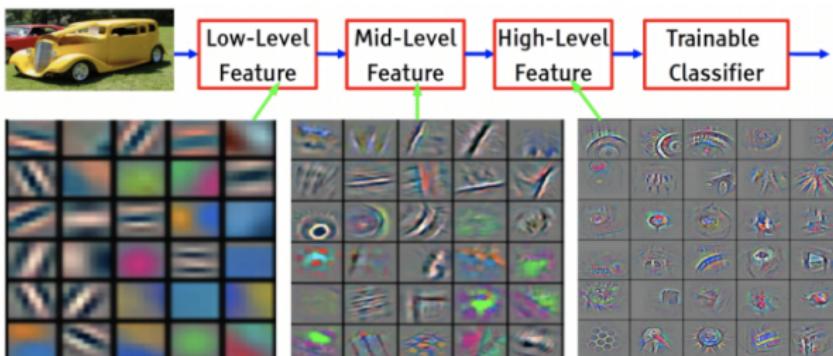


Figure: Réseau de neurones convolutionnel 2D

- Extraction de patterns locaux avec des filtres.
 - Découverte automatique des patterns à chercher.
- ① Découverte des orientations des traits et couleurs.
② Découverte de contours.

Introduction

Réseau convolutionnel

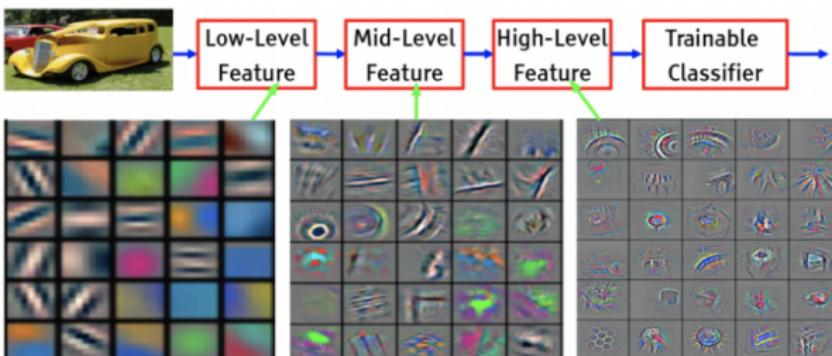


Figure: Réseau de neurones convolutionnel 2D

- Extraction de patterns locaux avec des filtres.
 - Découverte automatique des patterns à chercher.
- ① Découverte des orientations des traits et couleurs.
 - ② Découverte de contours.
 - ③ Découverte d'objets.

Introduction

Trois types de calculs

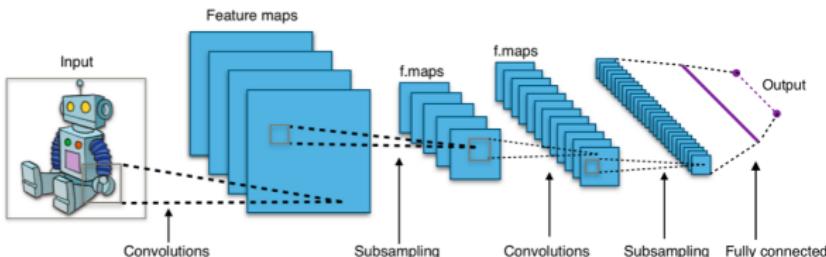


Figure: Réseau de neurones convolutionnel 2D

Convolutions : Extraction des caractéristiques locales.

Trois types de calculs

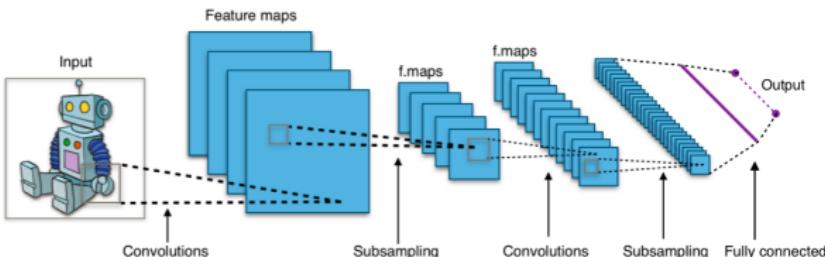


Figure: Réseau de neurones convolutionnel 2D

Convolutions : Extraction des caractéristiques locales.

Pooling : Invariance par translation et réduction de la taille des images.

Trois types de calculs

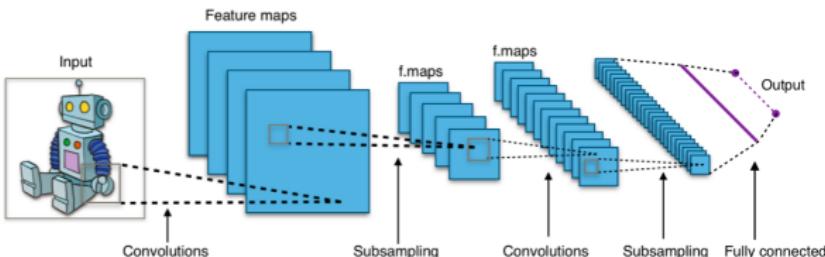


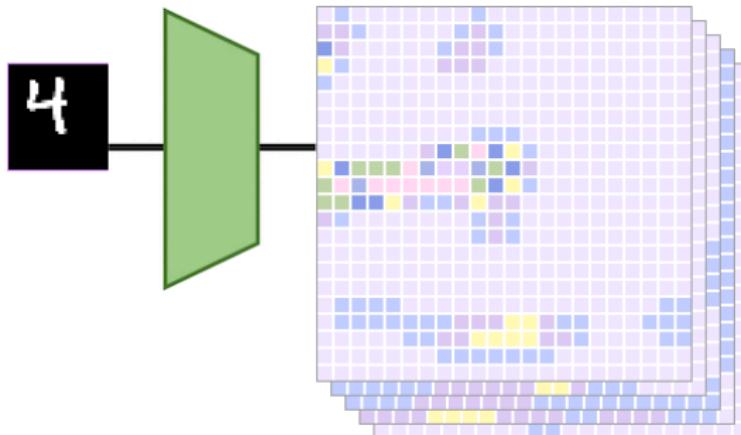
Figure: Réseau de neurones convolutionnel 2D

Convolutions : Extraction des caractéristiques locales.

Pooling : Invariance par translation et réduction de la taille des images.

Activation ReLU : Non-linéarité sans saturation.

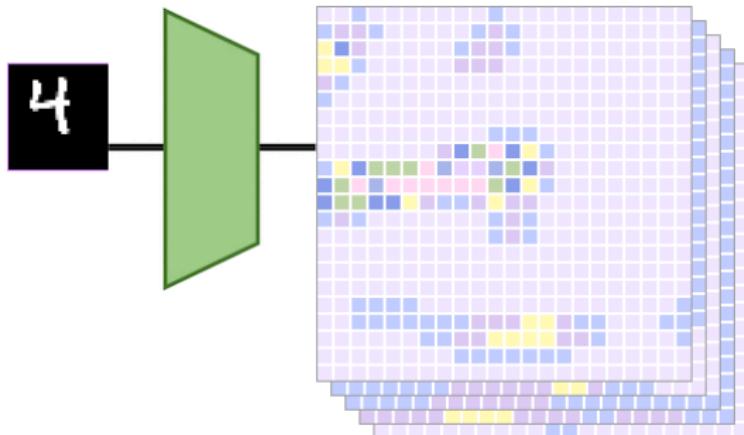
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

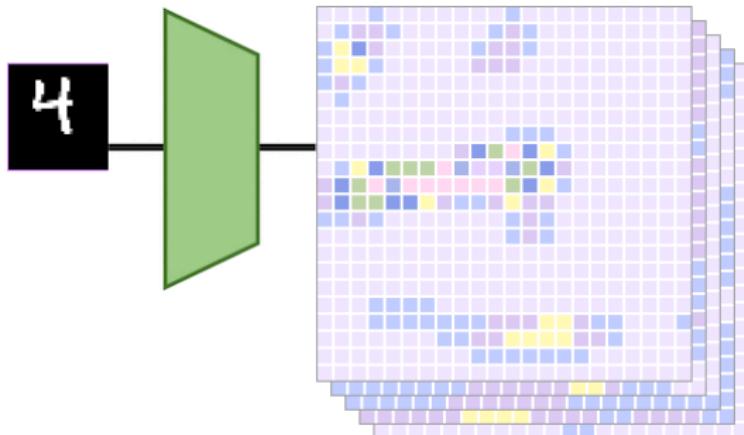
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

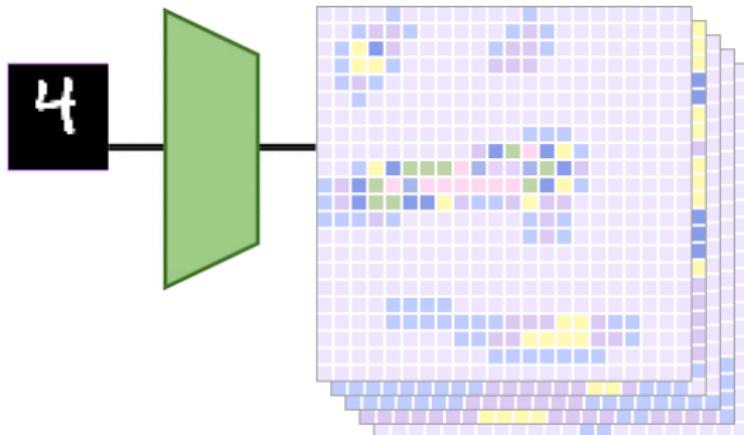
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

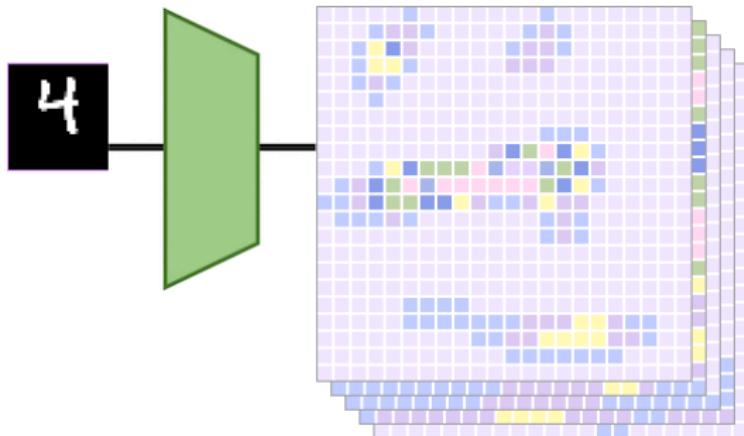
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

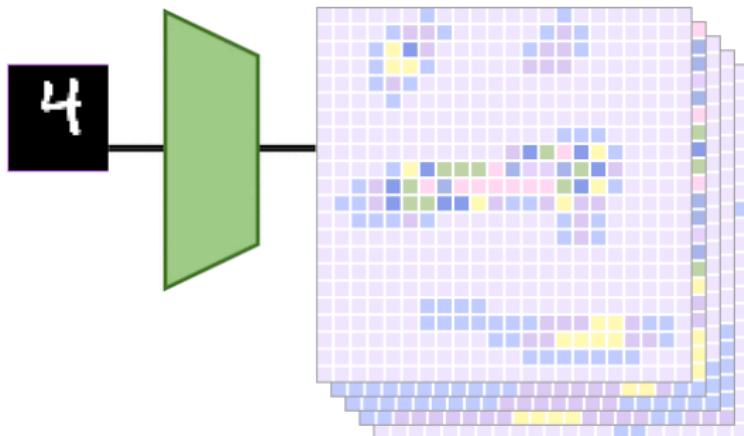
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution

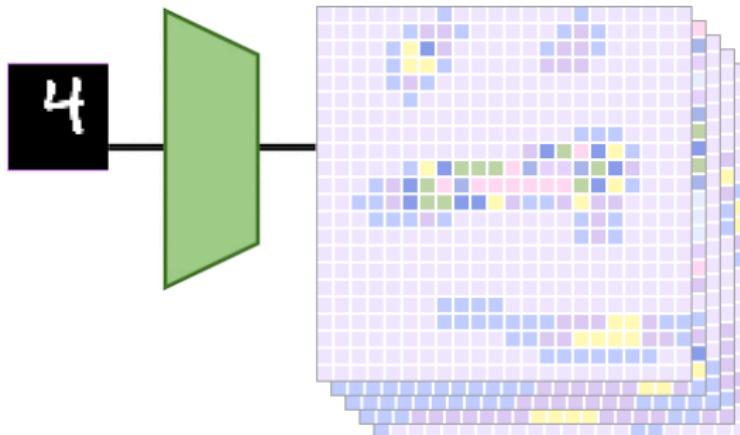


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

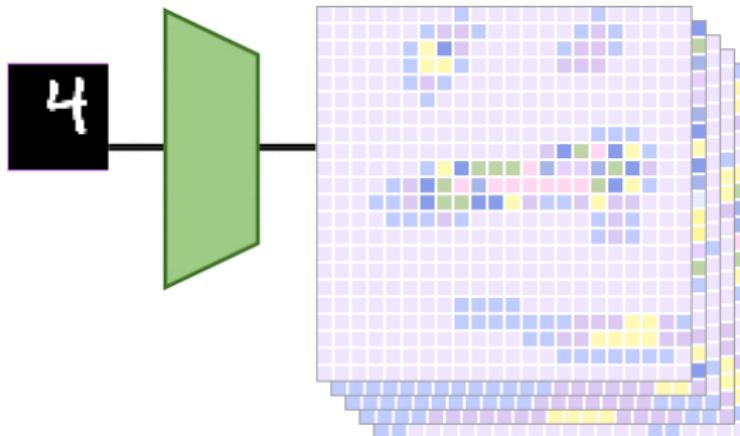
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

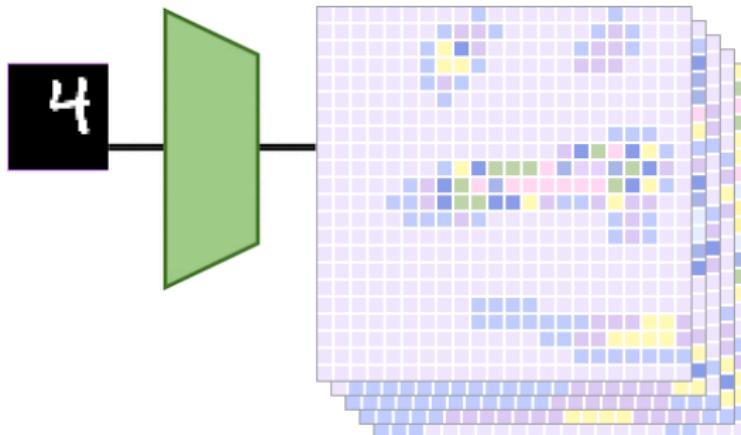
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

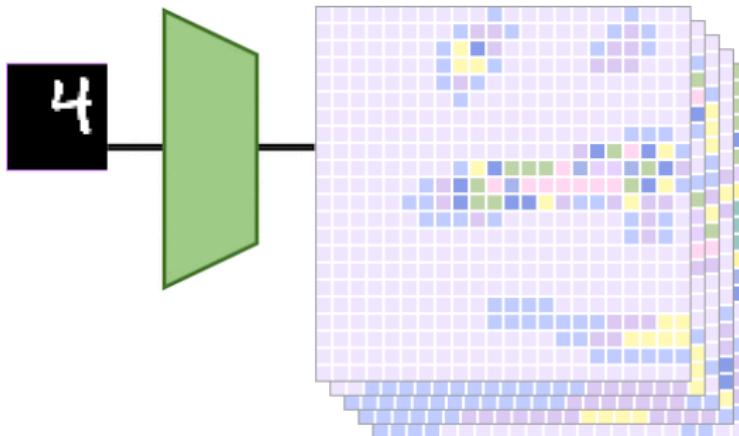
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

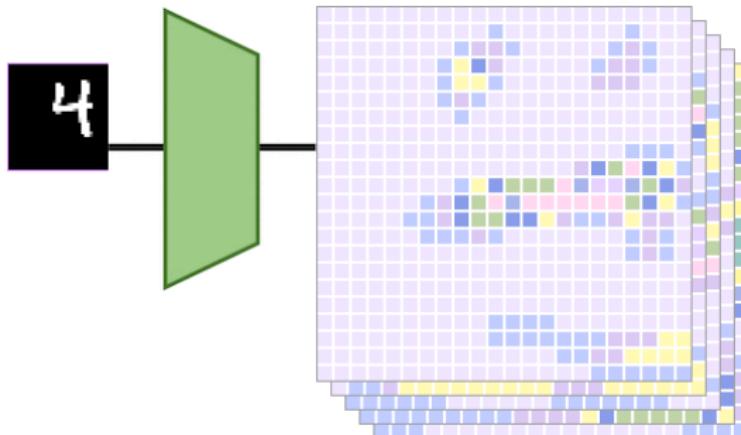
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

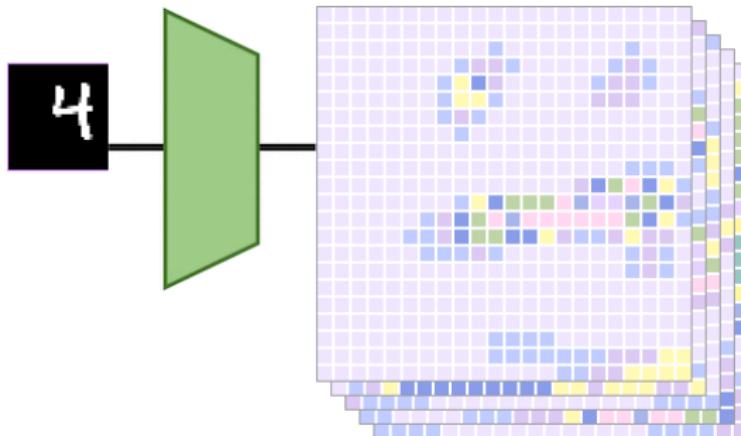
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

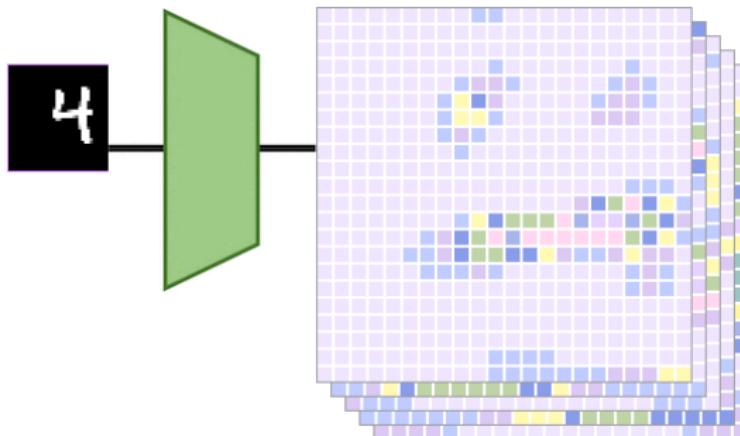
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

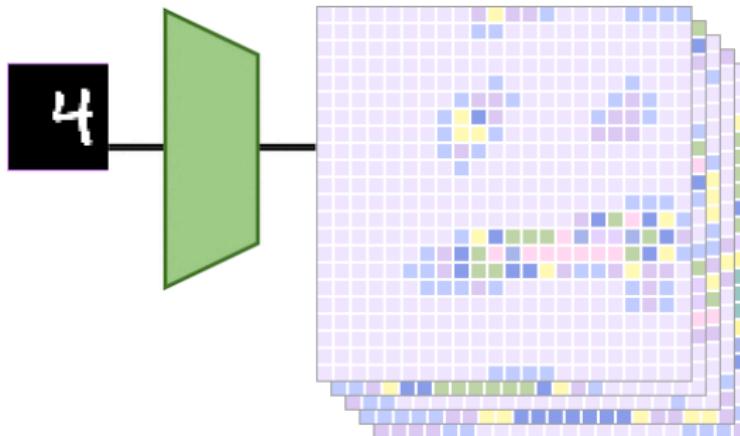
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

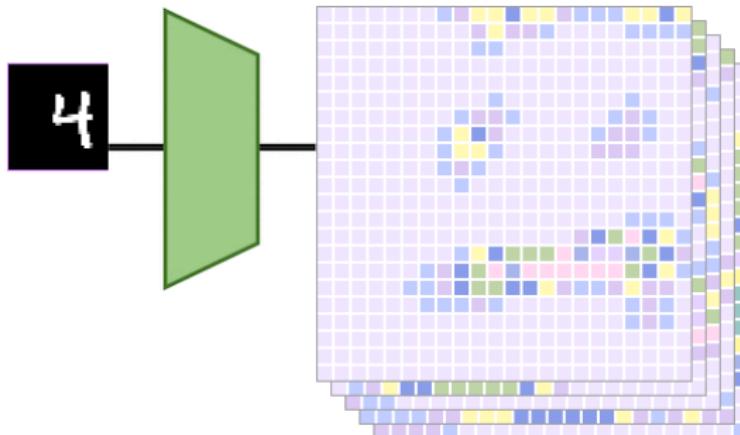
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution

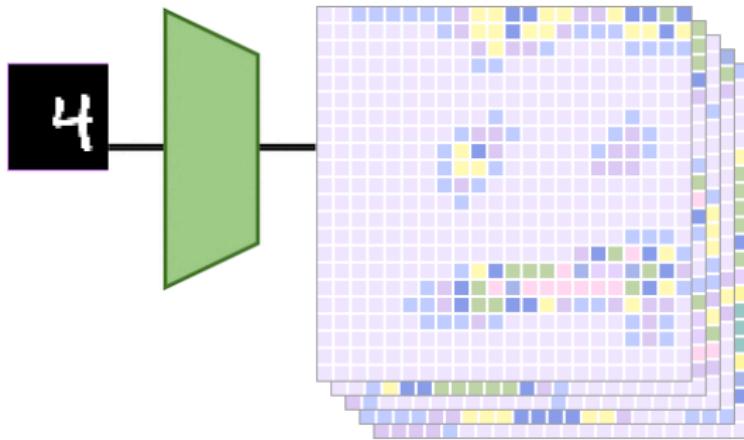


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

Illustration de la convolution

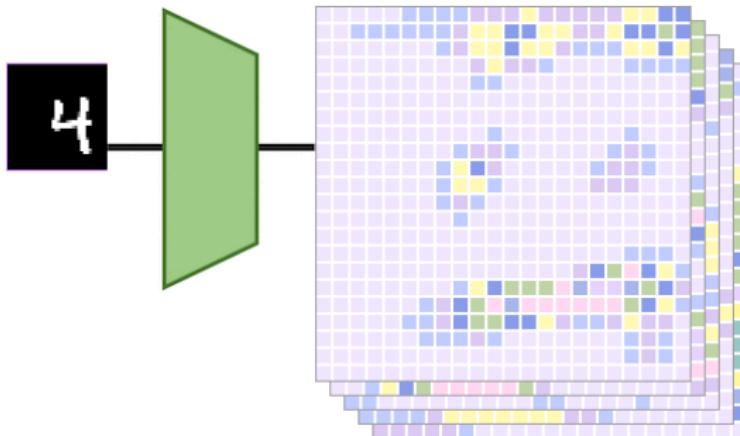


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

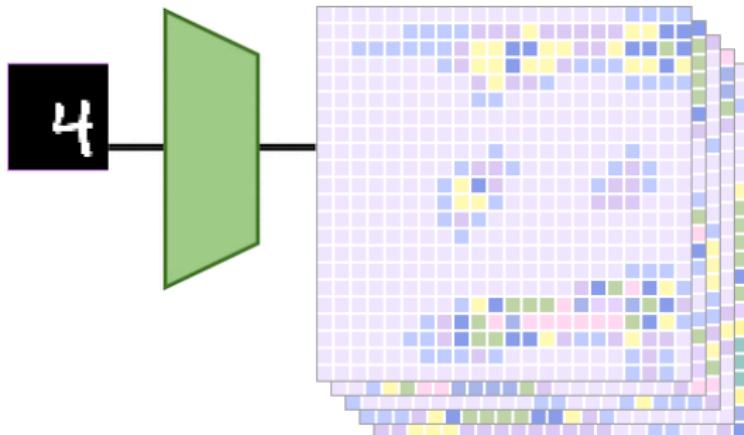
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

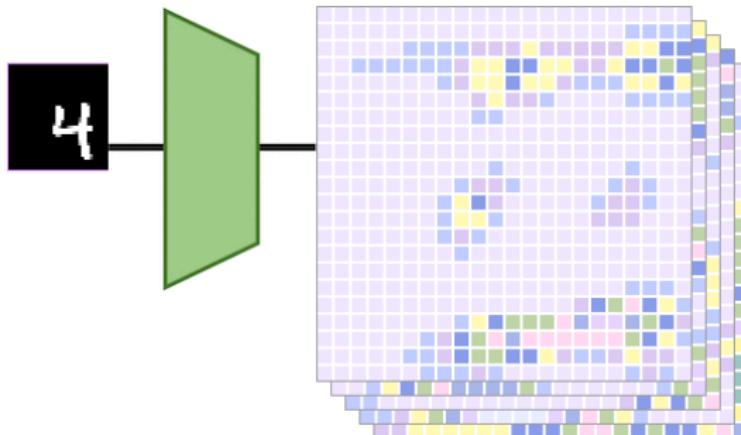
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution

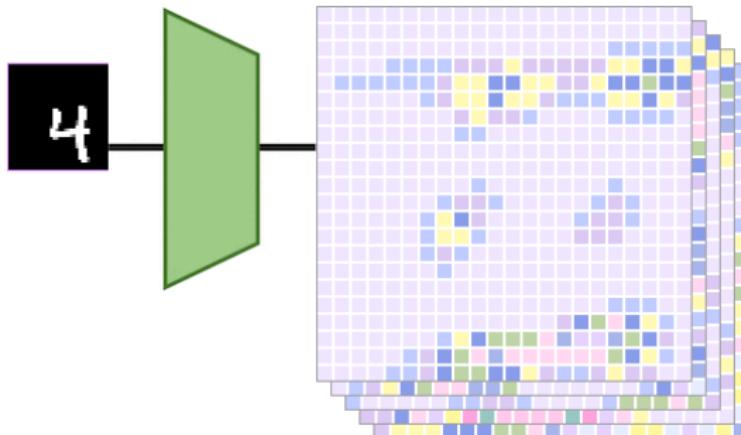


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

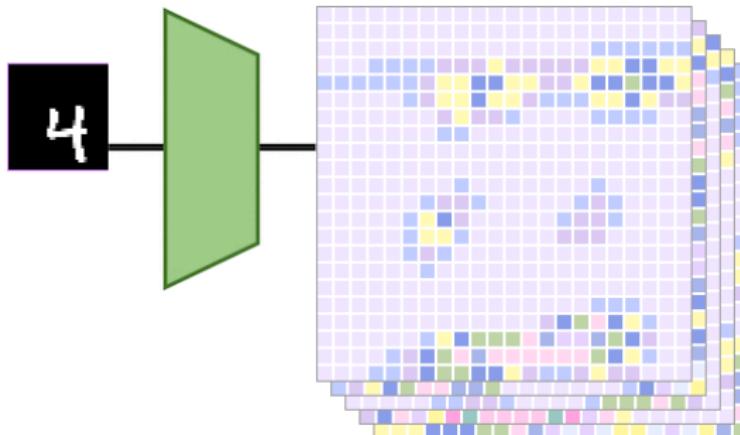
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

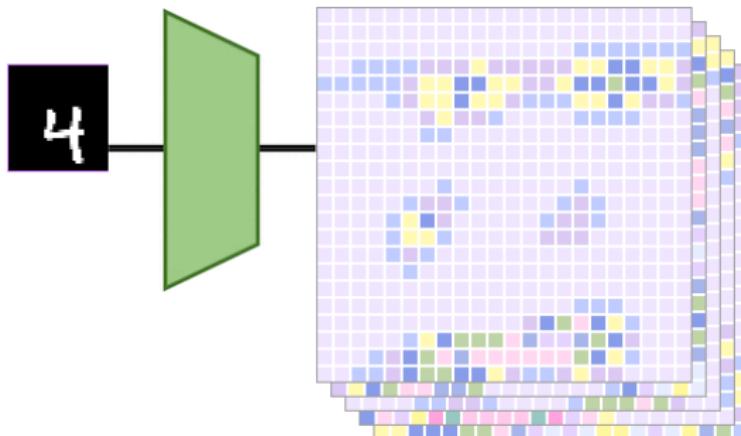
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

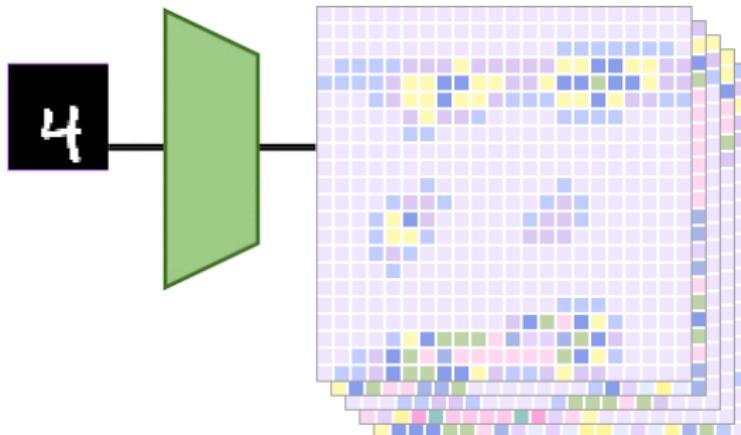
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

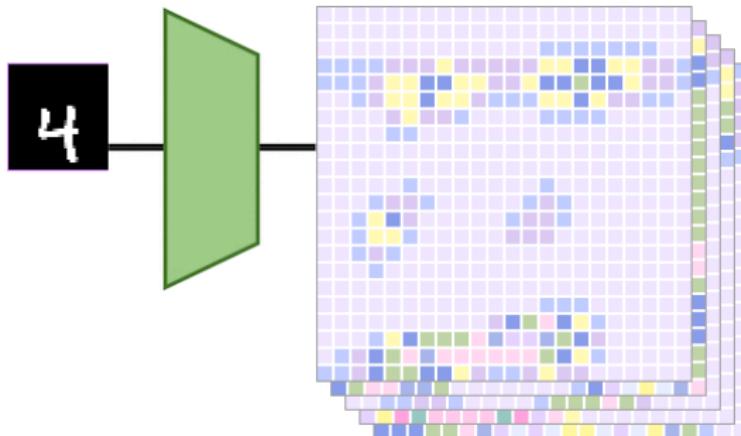
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

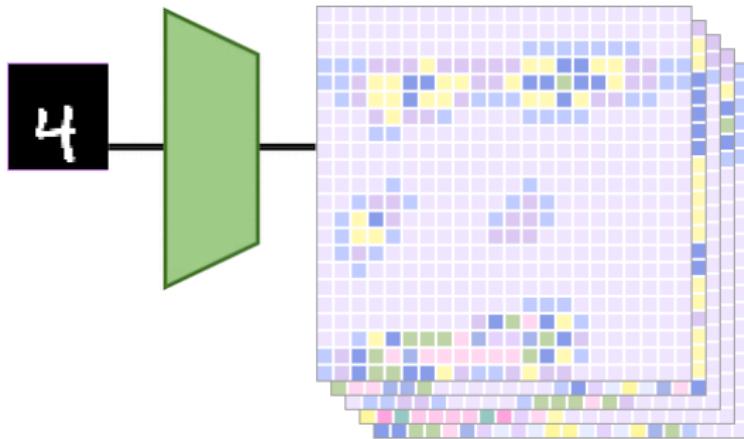
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifield.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

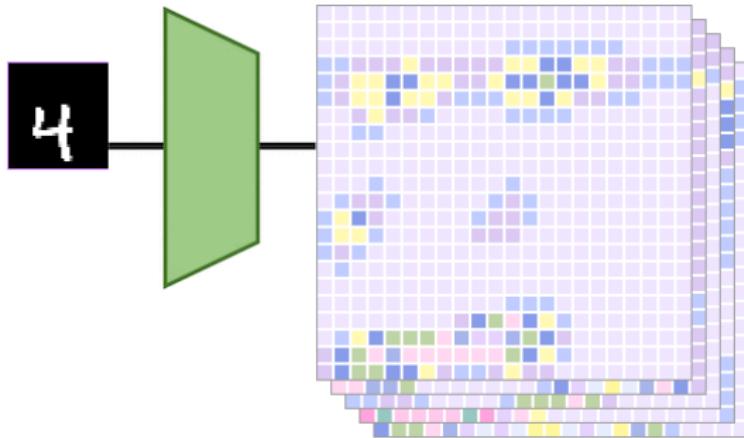
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

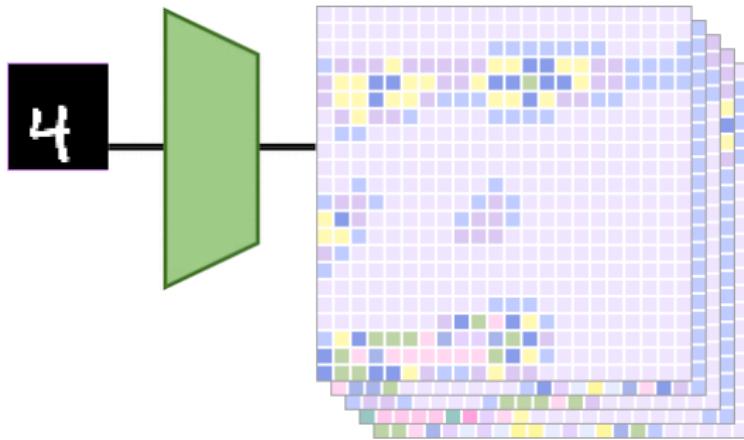
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chriswolfvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution

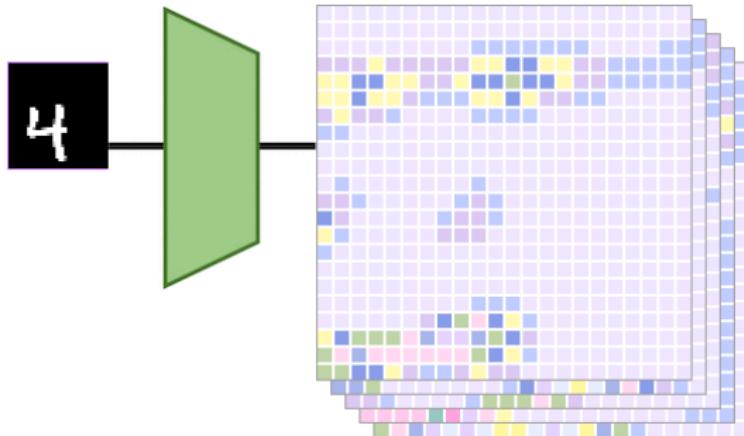


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

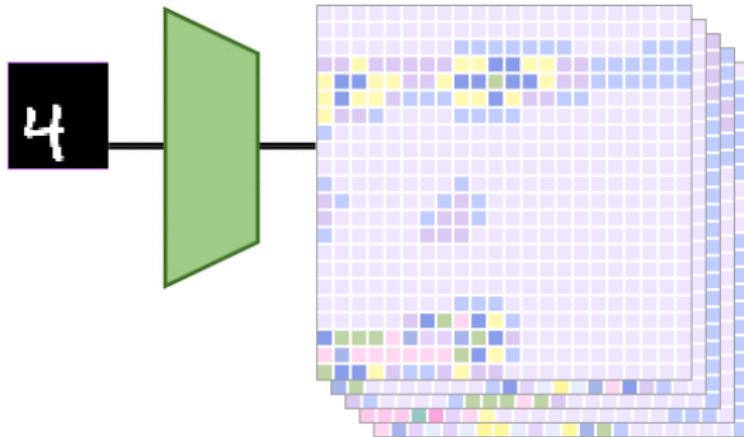
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

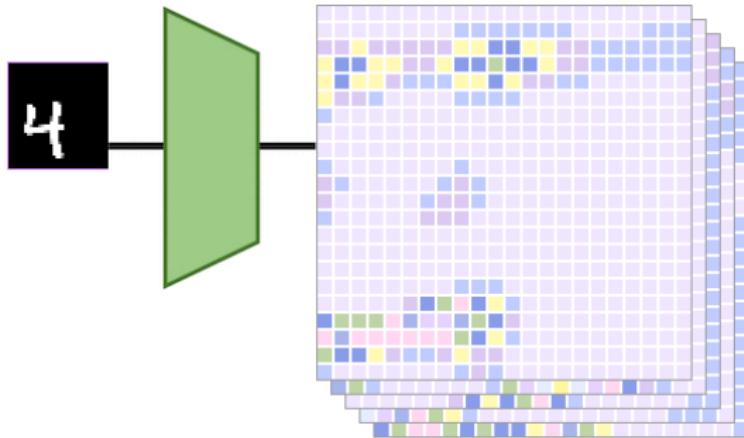
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

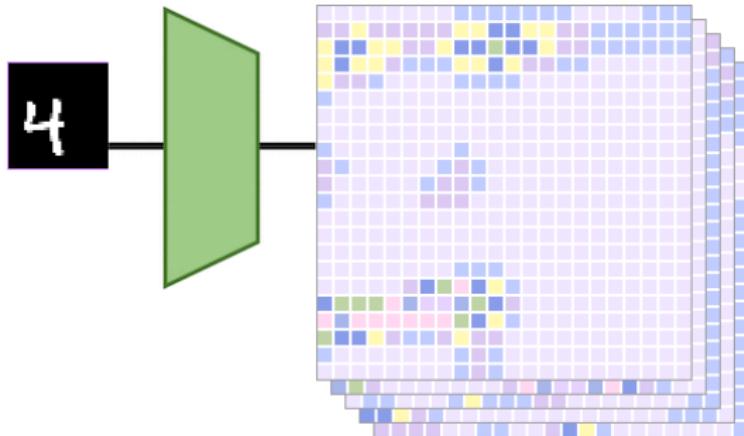
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

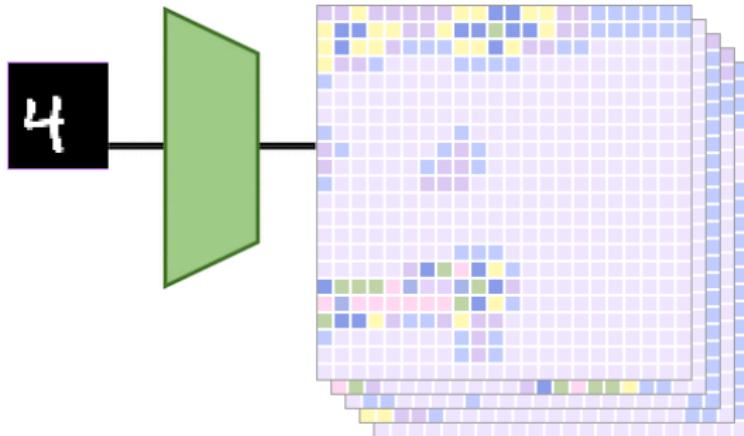
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

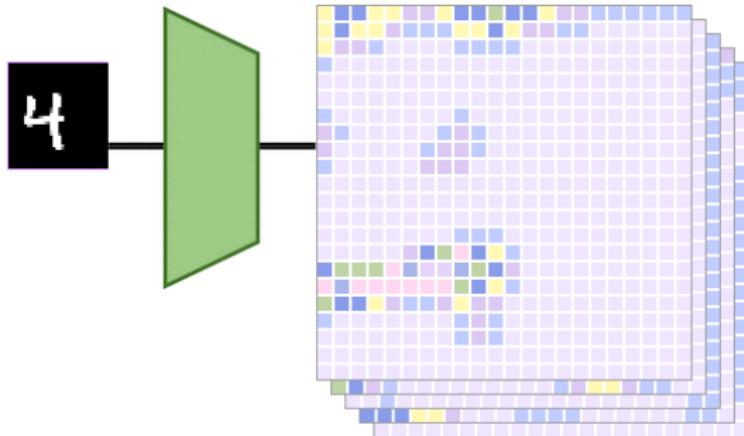
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution

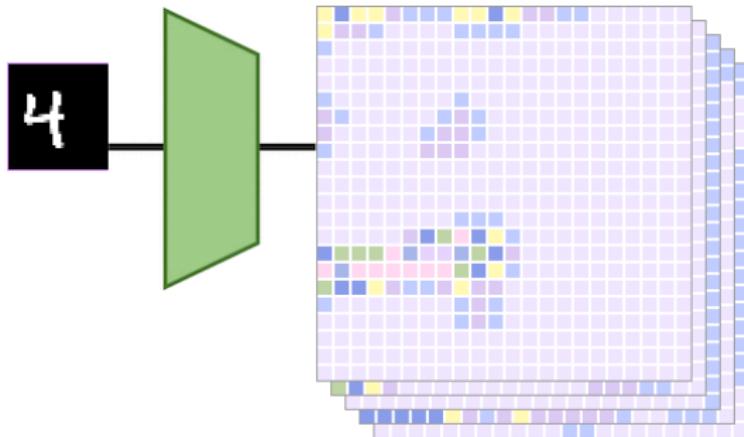


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

Illustration de la convolution

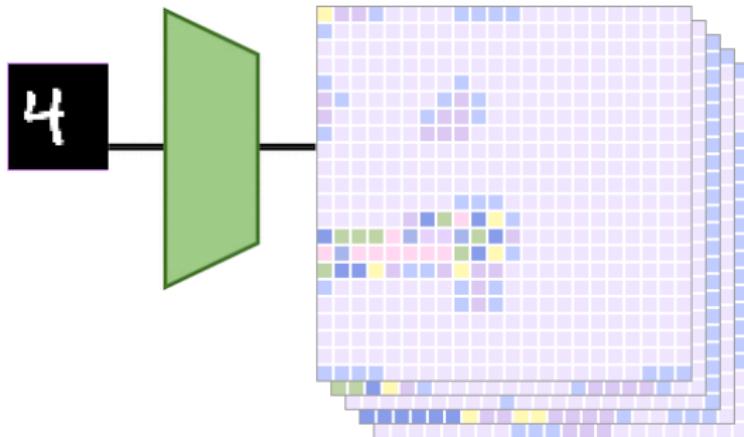


- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Introduction

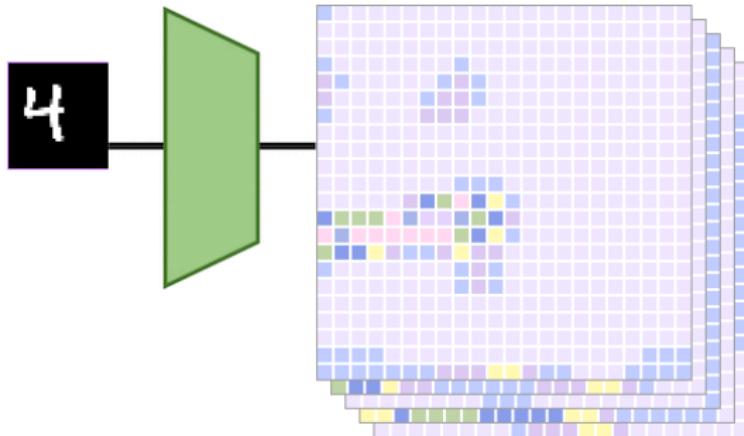
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

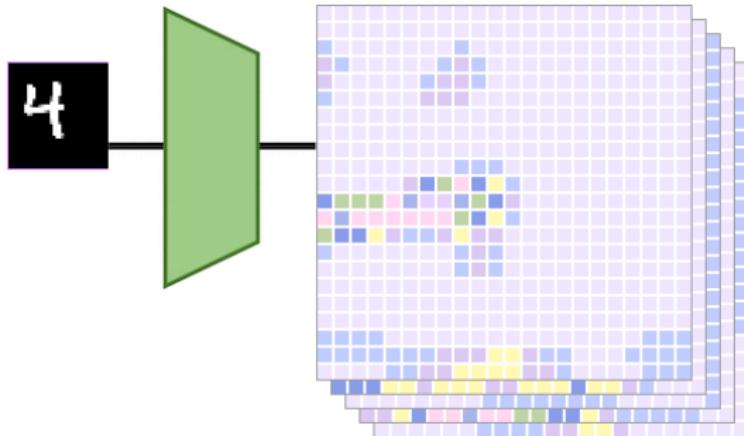
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

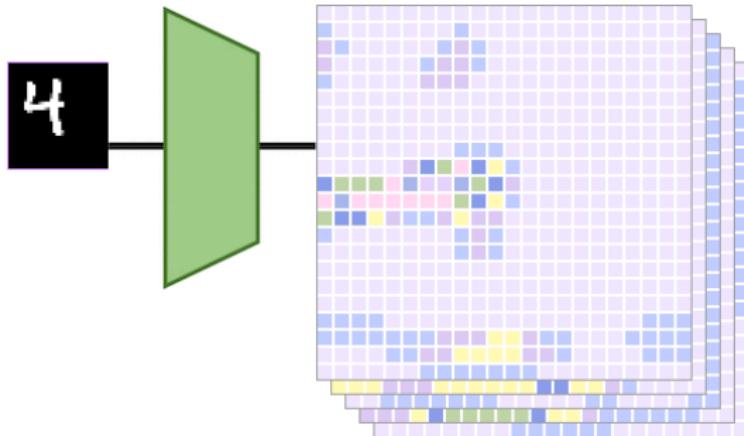
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

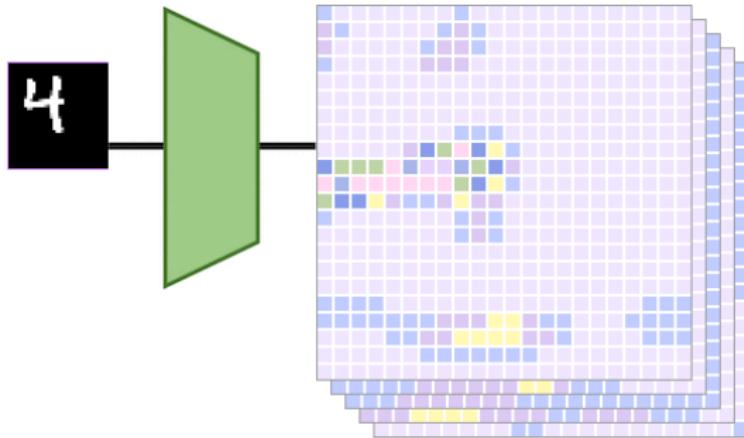
Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Illustration de la convolution



- Plusieurs patterns recherchés; une image -> plusieurs images.
- Décalage des patterns trouvés si on décale l'image.

<https://chrisholifordvision.medium.com/what-is-translation-equivariance-and-why-do-we-use-convolutions-to-get-it-6f18139d4c59>

Avantages/Limites



Figure: Invariances CNN.

Avantages:

- Equivariance par translation.

Désavantages:

Avantages/Limites



Figure: Invariances CNN.

Avantages:

- Equivariance par translation.

Désavantages:

- Variance par rotation.

Avantages/Limites



Figure: Invariances CNN.

Avantages:

- Equivariance par translation.

Désavantages:

- Variance par rotation.
- Variance par réflexion.

Avantages/Limites



Figure: Invariances CNN.

Avantages:

- Equivariance par translation.

Désavantages:

- Variance par rotation.
- Variance par réflexion.
- Non-encodage de la position.

Cross-corrélation

Appliquer d'une cross-corrélation entre g et f en u :

- 1D: $g_{filtre}(u) = \sum_{x=-\infty}^{\infty} f(x)g(x-u)$
- 2D: $g_{filtre}(u, v) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} f(x, y)g(x-u, y-v)$
- 3D: $g_{filtre}(u, v, w) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \sum_{z=-\infty}^{\infty} f(x, y, z)g(x-u, y-v, z-w)$
- etc...

Souvent, $\forall x \notin [-k, k], f(x) = 0$, avec k réel.

Masque convolution 1D

Masque convolution 1D

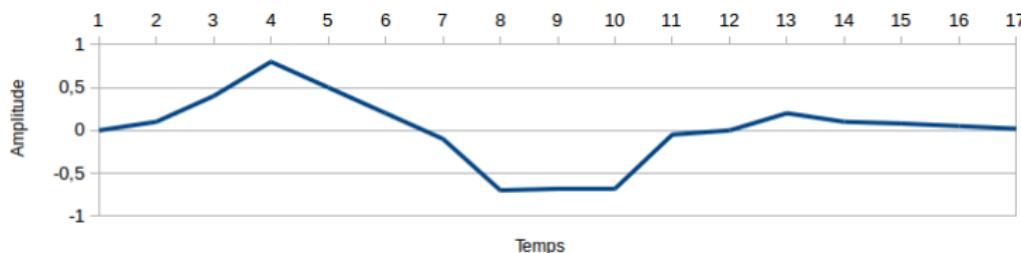


Figure: Serie de données F

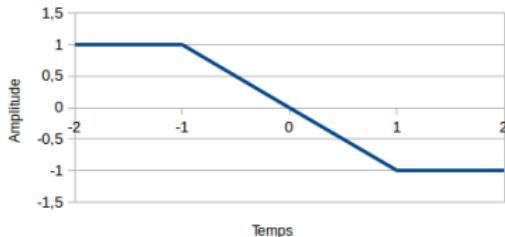


Figure: Masque de convolution G

Masque convolution 1D

Masque convolution 1D

0	0,1	0,4	0,8	0,5	0,2	-0,1	-0,7	-0,69	-0,68
-0,05	0	0,2	0,1	0,08	0,05	0,02	0	0	0

Table: Série temporelle F

1	1	0	-1	-1
---	---	---	----	----

Table: Masque de convolution G

- Appliquer le filtre G à F.

Masque convolution 1D

Masque convolution 1D

0	0,1	0,4	0,8	0,5	0,2	-0,1	-0,7	-0,69	-0,68
-0,05	0	0,2	0,1	0,08	0,05	0,02	0	0	0

Table: Série temporelle F

1	1	0	-1	-1
---	---	---	----	----

Table: Masque de convolution G

- Appliquer le filtre G à F.
- Masque=filtre=noyau de convolution.

Masque convolution 1D

Masque convolution 1D

0	0,1	0,4	0,8	0,5	0,2	-0,1	-0,7	-0,69	-0,68
-0,05	0	0,2	0,1	0,08	0,05	0,02	0	0	0

Table: Série temporelle F

1	1	0	-1	-1
---	---	---	----	----

Table: Masque de convolution G

- Appliquer le filtre G à F.
- Masque=filtre=noyau de convolution.
- Rechercher le **pattern** correspondant à G.

Masque convolution 1D

Appliquer masque convolution 1D



1	1	0	-1	-1
---	---	---	----	----

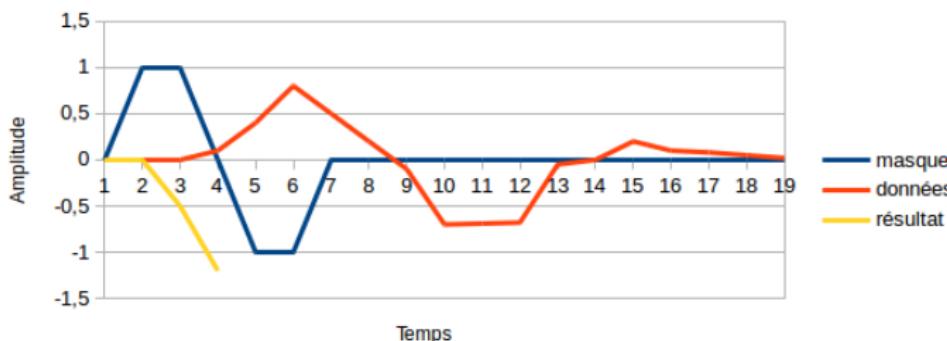
0	0,1	0,4	0,8	0,5	0,2	-0,1	-0,7	-0,69	-0,68
-0,05	0	0,2	0,1	0,08	0,05	0,02	0	0	0

$$1*0+1*0+0*0-1*0,1-1*0,4=-0,5$$

-0,5								

Masque convolution 1D

Appliquer masque convolution 1D



1	1	0	-1	-1
0	0,1	0,4	0,8	0,5
-0,05	0	0,2	0,1	0,08

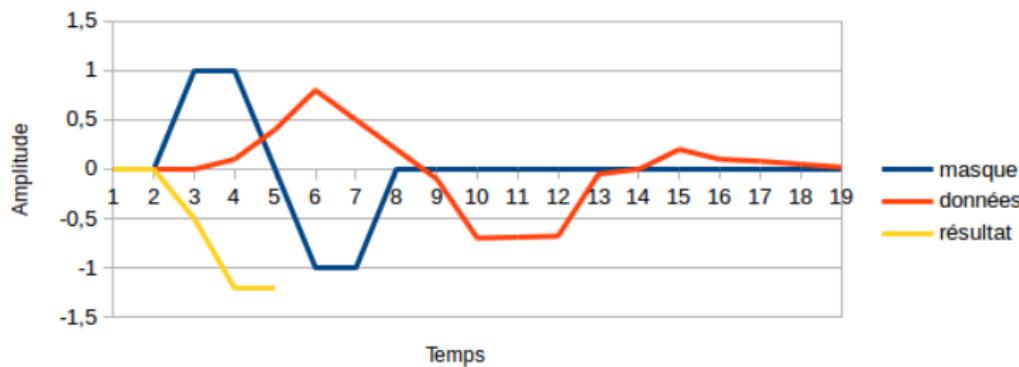
0,2	-0,1	-0,7	-0,69	-0,68
0,05	0,02	0	0	0

$$1*0+1*0+0*0,1-1*0,4-1*0,8=-1,2$$

-0,5	-1,2						

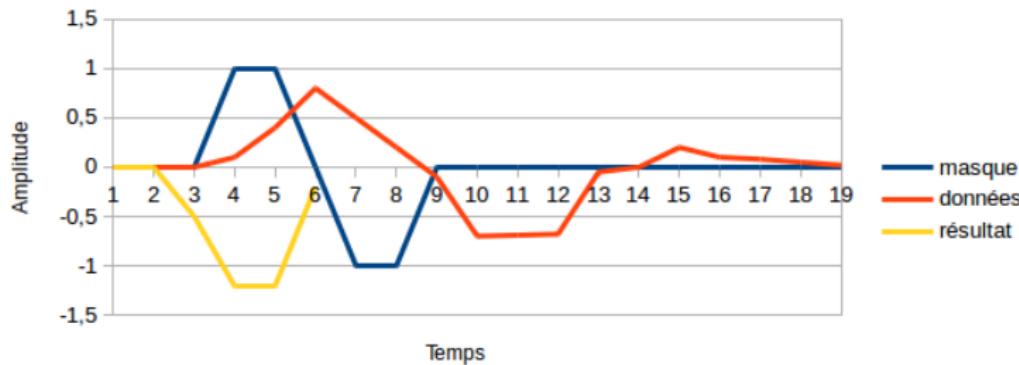
Masque convolution 1D

Appliquer masque convolution 1D



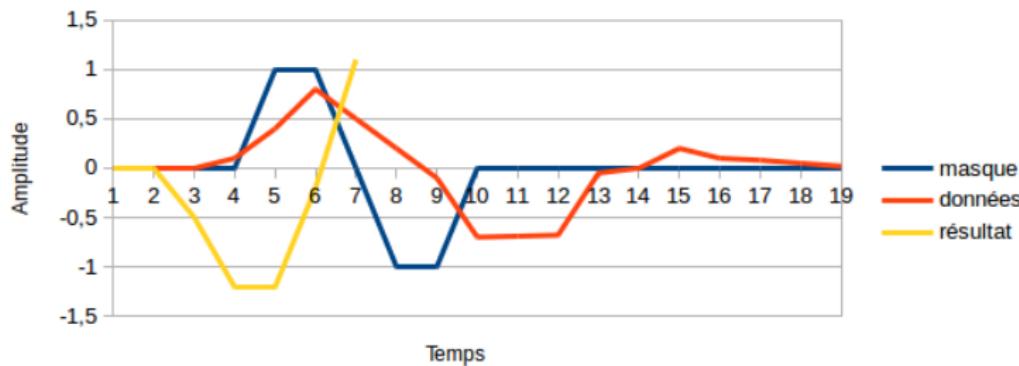
Masque convolution 1D

Appliquer masque convolution 1D



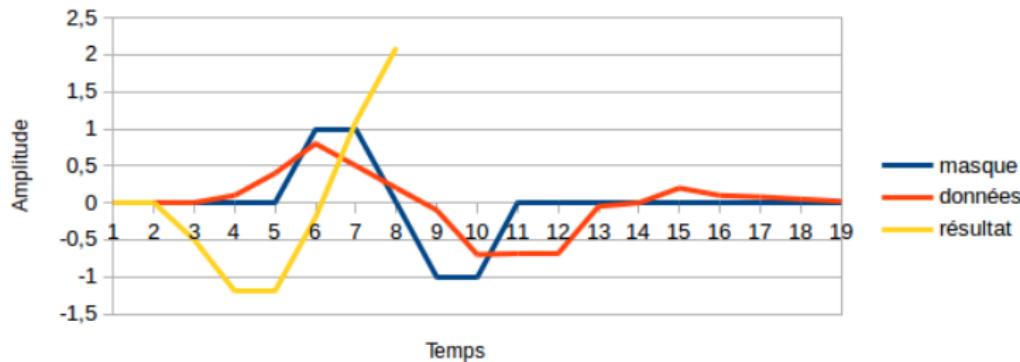
Masque convolution 1D

Appliquer masque convolution 1D



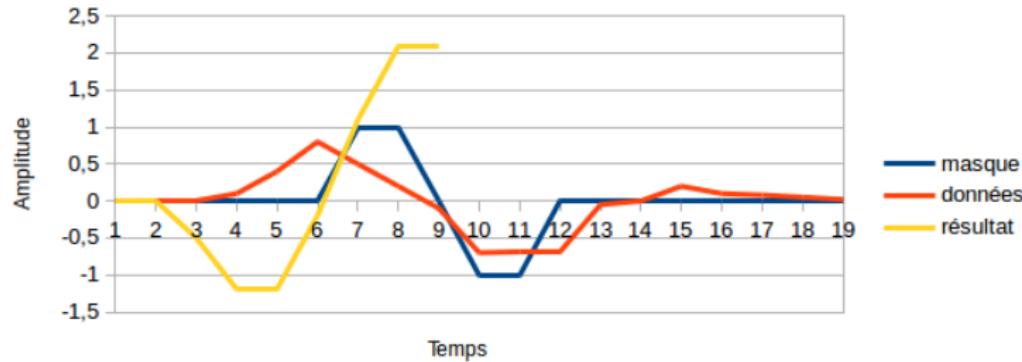
Masque convolution 1D

Appliquer masque convolution 1D



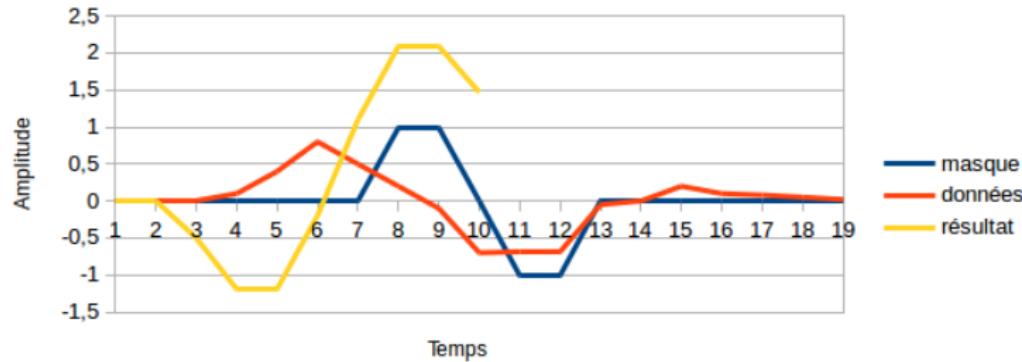
Masque convolution 1D

Appliquer masque convolution 1D



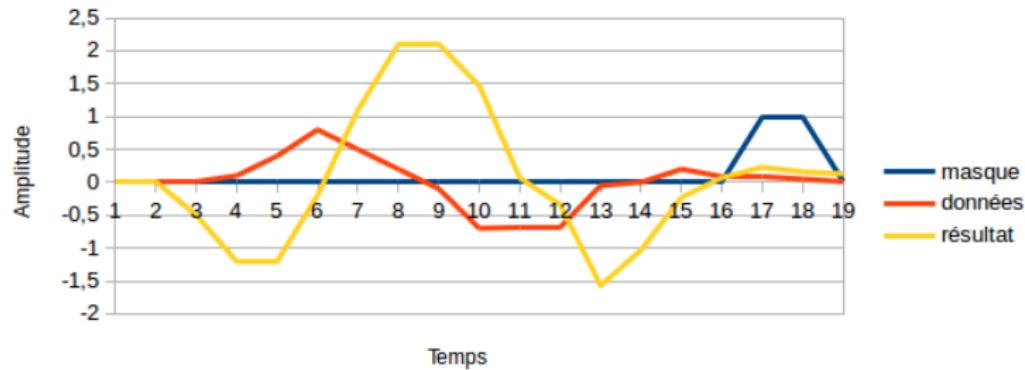
Masque convolution 1D

Appliquer masque convolution 1D



Masque convolution 1D

Appliquer masque convolution 1D



Masque convolution 1D

Appliquer masque convolution 1D

0	0,1	0,4	0,8	0,5	0,2	-0,1	-0,7	-0,69	-0,68
-0,05	0	0,2	0,1	0,08	0,05	0,02	0	0	0

-0,5	-1,2	-1,2	-0,2	1,1	2,1	2,09	1,47	0,07
-0,34	-1,57	-1,03	-0,23	0,07	0,23	0,16	0,13	0,02

Masque convolution 2D

Masque convolution 2D

3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

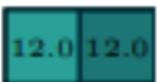
12.0

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3_0	2_1	1_2	0
0	0_2	1_2	3_0	1
3	1_0	2_1	2_2	3
2	0	0	2	2
2	0	0	0	1



Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2 ₀	1 ₁	0 ₂
0	0	1 ₂	3 ₂	1 ₀
3	1	2 ₀	2 ₁	3 ₂
2	0	0	2	2
2	0	0	0	1

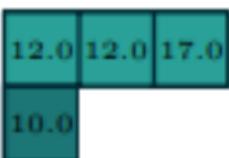
12.0	12.0	17.0
------	------	------

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2	1	0
0 ₀	0 ₁	1 ₂	3	1
3 ₂	1 ₂	2 ₀	2	3
2 ₀	0 ₁	0 ₂	2	2
2	0	0	0	1



Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2	1	0
0	0 ₀	1 ₁	3 ₂	1
3	1 ₂	2 ₂	2 ₀	3
2	0 ₀	0 ₁	2 ₂	2
2	0	0	0	1

12.0	12.0	17.0
10.0	17.0	

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2	1	0
0	0	1_0	3_1	1_2
3	1	2_2	2_2	3_0
2	0	0_0	2_1	2_2
2	0	0	0	1

12.0	12.0	17.0
10.0	17.0	19.0

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2	1	0
0	0	1	3	1
3 ₀	1 ₁	2 ₂	2	3
2 ₂	0 ₂	0 ₀	2	2
2 ₀	0 ₁	0 ₂	0	1

12.0	12.0	17.0
10.0	17.0	19.0
9.0		

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

3	3	2	1	0
0	0	1	3	1
3	1 ₀	2 ₁	2 ₂	3
2	0 ₂	0 ₂	2 ₀	2
2	0 ₀	0 ₁	0 ₂	1

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

Masque convolution 2D

Masque convolution 2D

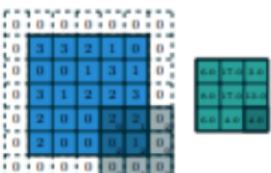
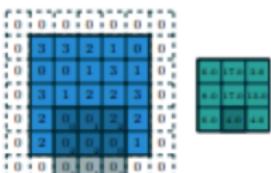
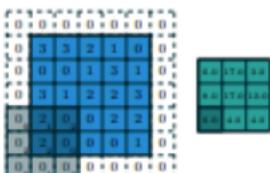
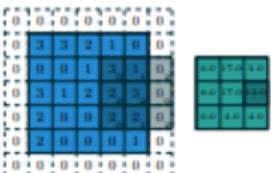
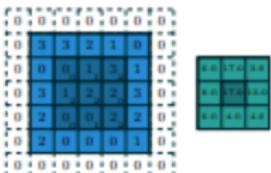
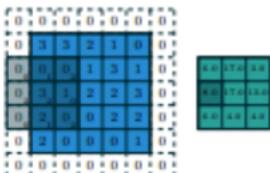
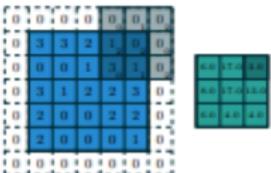
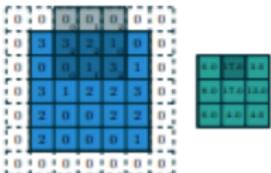
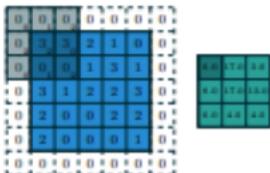
3	3	2	1	0
0	0	1	3	1
3	1	2 ₀	2 ₁	3 ₂
2	0	0 ₂	2 ₂	2 ₀
2	0	0 ₀	0 ₁	1 ₂

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning.

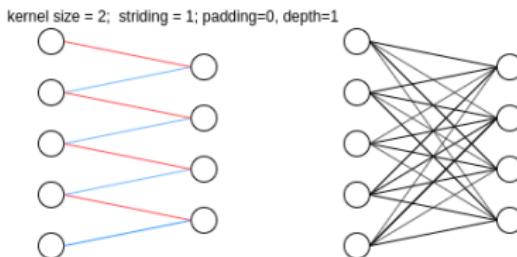
Masque convolution 2D

Padding et striding



Masque convolution 2D

Calculer la taille de la couche de sortie



F : taille du filtre de convolution.

W : taille des données d'entrées.

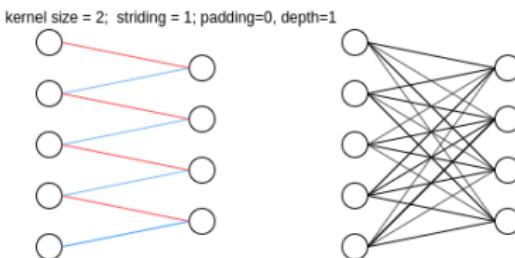
P : taille du padding.

S : taille du striding.

- Taille de la couche de sortie: $(W - F + 2P)/S + 1$.

Masque convolution 2D

Calculer la taille de la couche de sortie



F : taille du filtre de convolution.

W : taille des données d'entrées.

P : taille du padding.

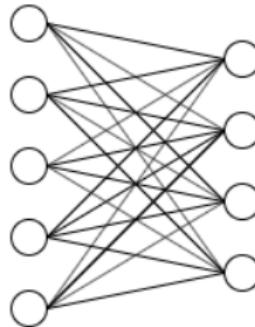
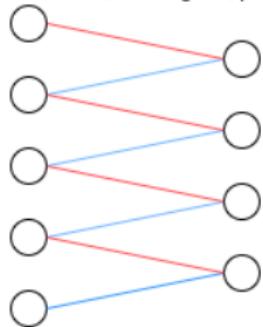
S : taille du striding.

- Taille de la couche de sortie: $(W - F + 2P)/S + 1$.
- Ici: $(5 - 2 + 0 * P)/1 + 1 = 4$.

Masque convolution 2D

Linéarité

kernel size = 2; striding = 1; padding=0, depth=1

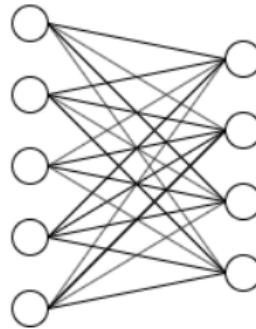
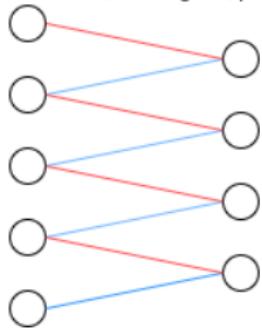


- Opération linéaire.

Masque convolution 2D

Linéarité

kernel size = 2; striding = 1; padding=0, depth=1

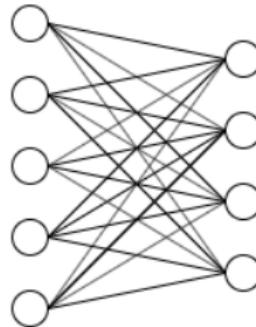
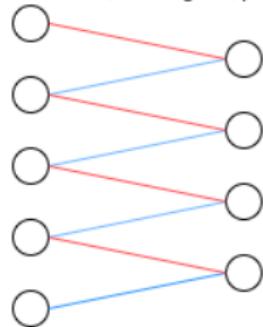


- Opération linéaire.
- Poids à 0.

Masque convolution 2D

Linéarité

kernel size = 2; striding = 1; padding=0, depth=1

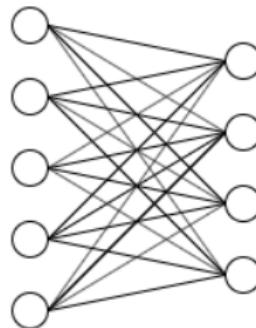
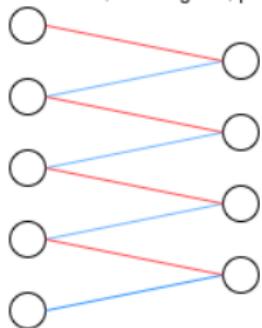


- Opération linéaire.
- Poids à 0.
- Poids partagés.

Masque convolution 2D

Linéarité

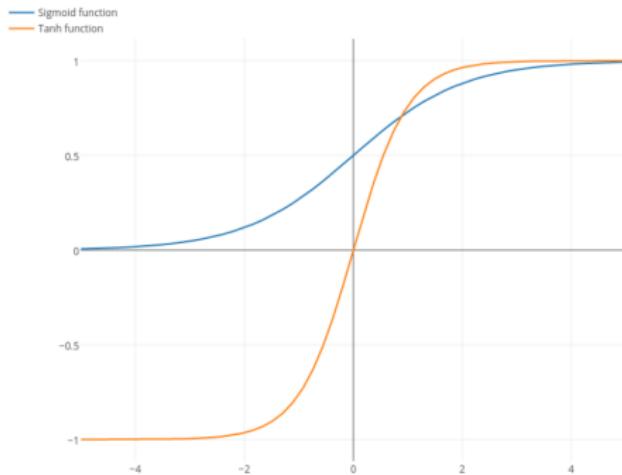
kernel size = 2; striding = 1; padding=0, depth=1



- Opération linéaire.
- Poids à 0.
- Poids partagés.
- Localité de l'information.

Masque convolution 2D

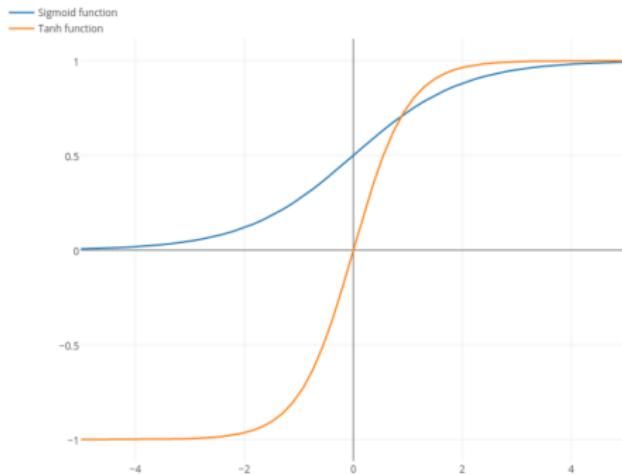
Sigmoid et Tanh



- Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$.

Masque convolution 2D

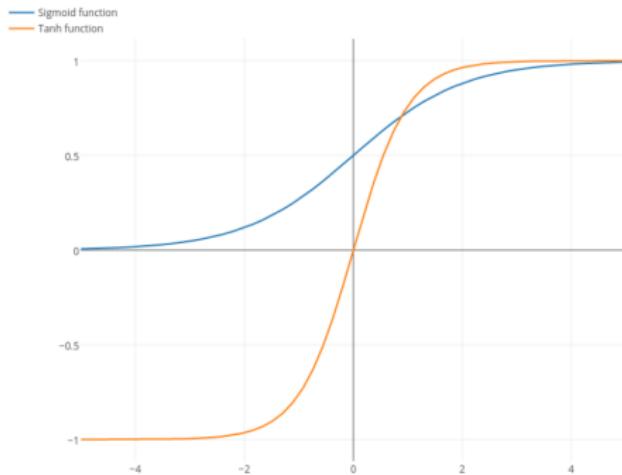
Sigmoid et Tanh



- Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$.
- $f'(x) = f(x)(1 - f(x))$.

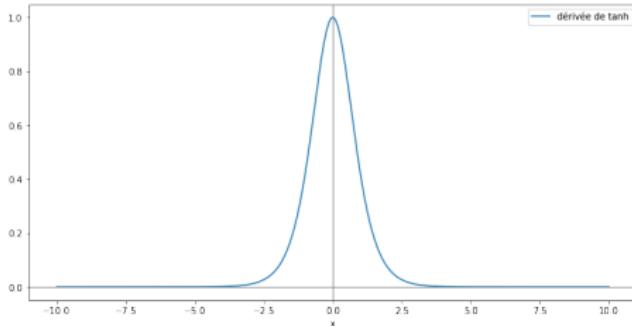
Masque convolution 2D

Sigmoid et Tanh



- Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$.
- $f'(x) = f(x)(1 - f(x))$.
- Saturation du gradient.

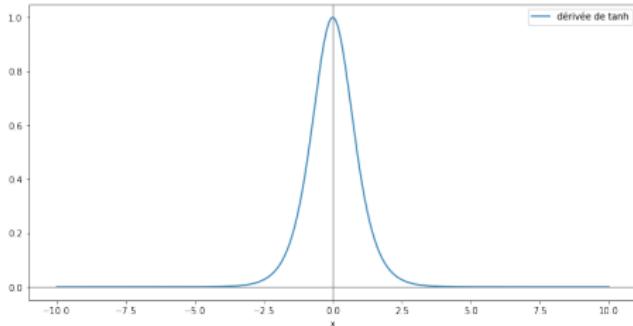
Vanishing gradient



- Réseau profond.

Masque convolution 2D

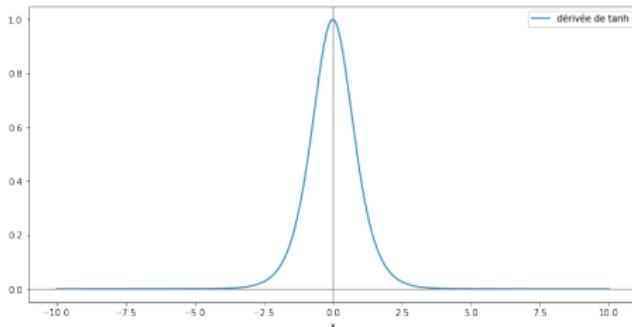
Vanishing gradient



- Réseau profond.
- Multiplication de valeurs entre $[0; 1]$ lors de la back-propagation.

Masque convolution 2D

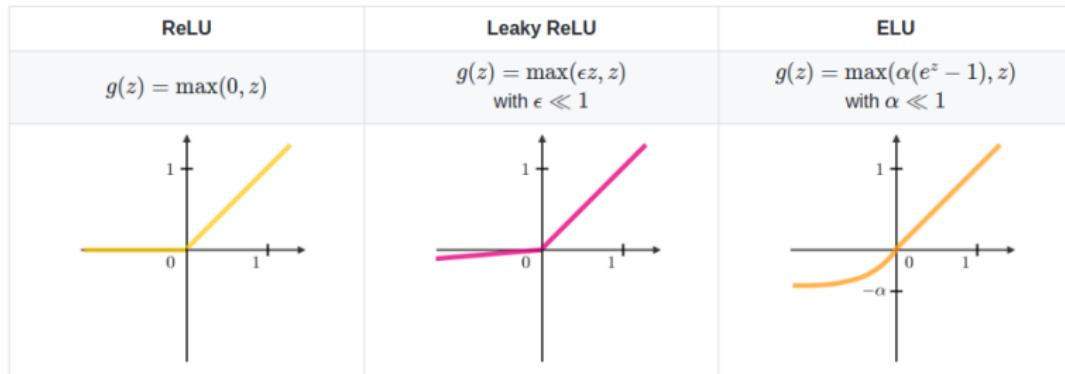
Vanishing gradient



- Réseau profond.
- Multiplication de valeurs entre $[0; 1]$ lors de la back-propagation.
- Gradient proche de 0 pour les premières couches.

Masque convolution 2D

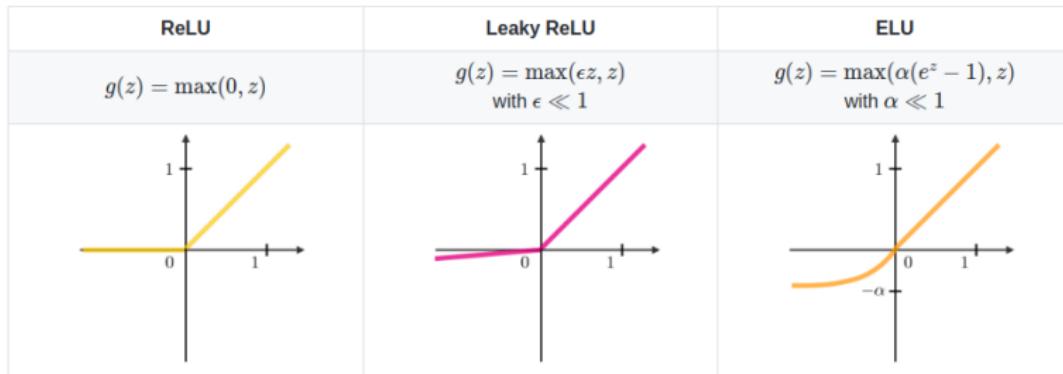
Fonctions d'activations



- ReLU ne sature pas: gradient constant.

Masque convolution 2D

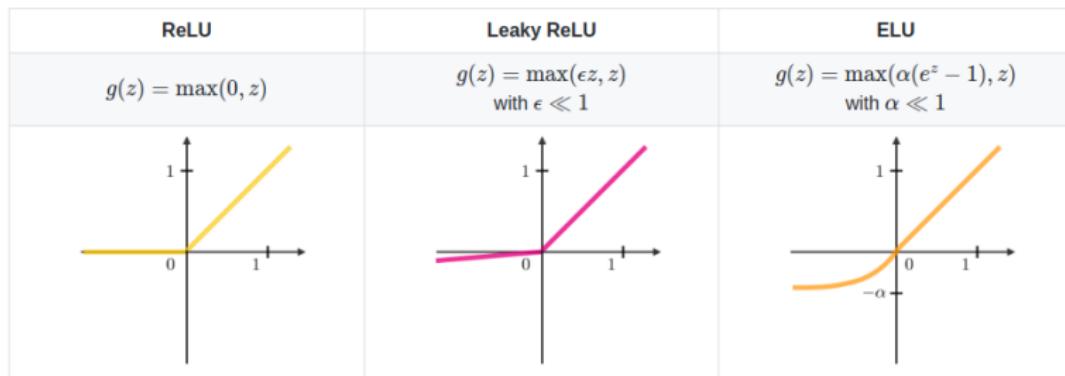
Fonctions d'activations



- ReLU ne sature pas: gradient constant.
- *dead neurons*: gradient inexistant si valeurs toujours négatives.

Masque convolution 2D

Fonctions d'activations

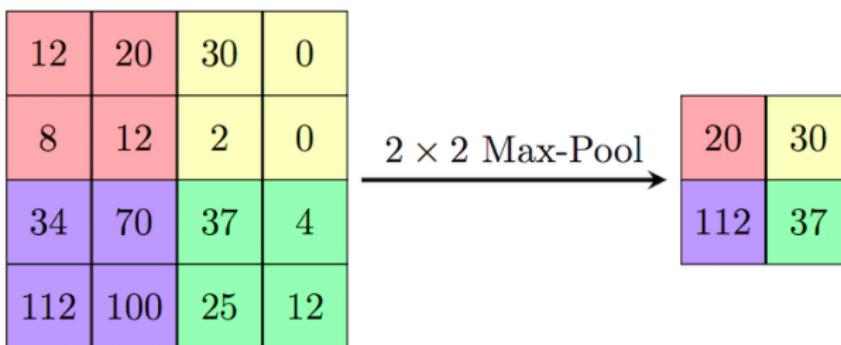


- ReLU ne sature pas: gradient constant.
- *dead neurons*: gradient inexistant si valeurs toujours négatives.
- LeakyReLU et ELU corrigent cela.

Masque convolution 2D

Max pooling

- Permet de réduire le nombre de neurones.
- Invariance par translation.



Masque convolution 2D

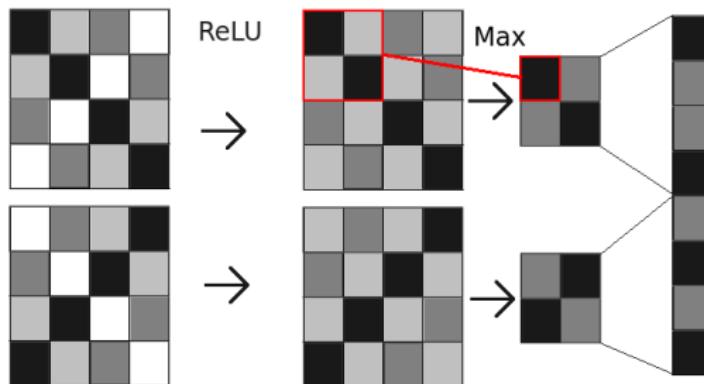
Exemple - étape 1 : convolution

The diagram illustrates a 2D convolution operation. It shows two input images (4x4 grids) being multiplied by two different 2x2 filters. The first multiplication results in a 4x4 output grid where each element is the sum of the overlapping elements from both inputs. The second multiplication results in a 4x4 output grid where each element is the product of the overlapping elements from both inputs.

- ➊ Deux filtres, donc deux images filtrées;
- ➋ $stride=1$, $padding=0$.

Masque convolution 2D

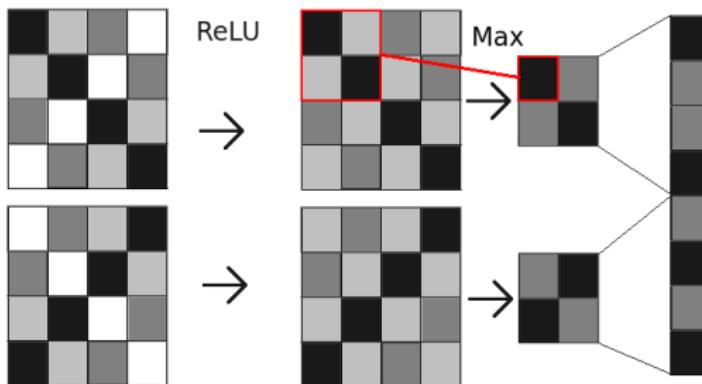
Exemple - étape 2 : ReLU + Max pooling



- Application de ReLU.

Masque convolution 2D

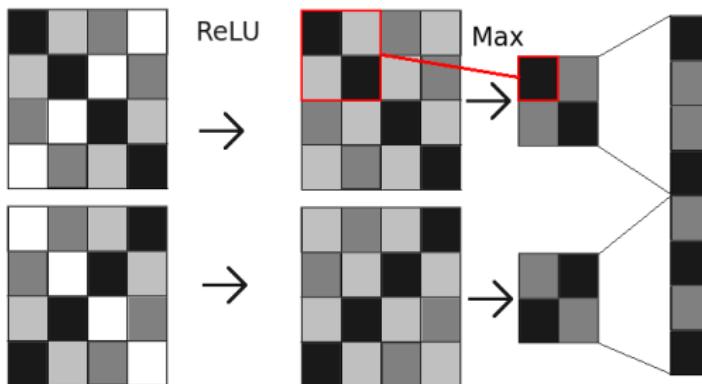
Exemple - étape 2 : ReLU + Max pooling



- Application de ReLU.
- Max-pooling.

Masque convolution 2D

Exemple - étape 2 : ReLU + Max pooling

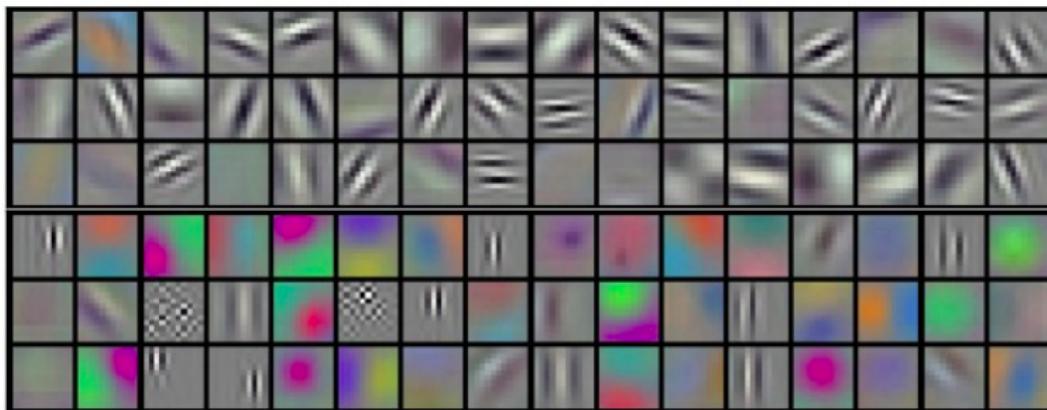


- Application de ReLU.
- Max-pooling.
- Mise à plat.
- Il a repéré 4 morceaux de diagonales.

Apprentissage des filtres

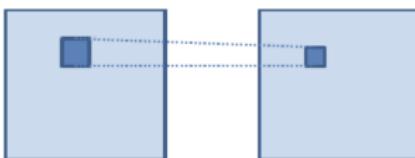
Avant : Définition manuelles des caractéristiques.

Deep learning : Apprentissage via back-propagation.

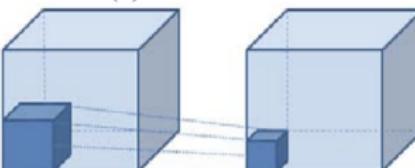


Masque convolution 2D

Convolutions 3D



(a) 2D convolution



(b) 3D convolution

3. Comparison of (a) 2D convolution and (b)

- On ajoute une nouvelle dimension spatiale.

<https://paperswithcode.com/method/3d-convolution>

Masque convolution 2D

Datasets 3D



- Ajout d'une dimension de volume (médecine).

Masque convolution 2D

Datasets 3D



- Ajout d'une dimension de volume (médecine).
- Vidéos: image 2D + temps.

Hypothèses utilisées pour biaiser notre réseau.

- Convolution 1D: patterns locaux 1D (Audio).

Hypothèses utilisées pour biaiser notre réseau.

- Convolution 1D: patterns locaux 1D (Audio).
- Convolution 2D: patterns locaux 2D (images).

Autres types de données: brève introduction

Hypothèses utilisées pour biaiser notre réseau.

- Convolution 1D: patterns locaux 1D (Audio).
- Convolution 2D: patterns locaux 2D (images).
- Convolution 3D: patterns locaux 3D (vidéo, volumes d'objets).

Autres types de données: brève introduction

Hypothèses utilisées pour biaiser notre réseau.

- Convolution 1D: patterns locaux 1D (Audio).
- Convolution 2D: patterns locaux 2D (images).
- Convolution 3D: patterns locaux 3D (vidéo, volumes d'objets).
- Biaiser notre architecture selon la structure des données.

Autres types de données: brève introduction

Hypothèses utilisées pour biaiser notre réseau.

- Convolution 1D: patterns locaux 1D (Audio).
- Convolution 2D: patterns locaux 2D (images).
- Convolution 3D: patterns locaux 3D (vidéo, volumes d'objets).
- Biaiser notre architecture selon la structure des données.
- D'autres types de structures pour d'autres données ? Des idées ?

Autres types de données: brève introduction

Séquences temporelles et réseaux récurrents

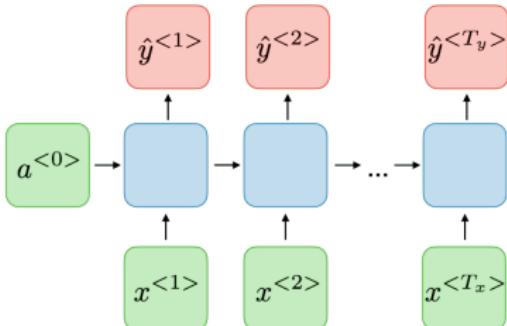


Figure: Illustration d'un RNN générique.

- Les données ont un ordre d'apparition avec une même structure.

Autres types de données: brève introduction

Séquences temporelles et réseaux récurrents

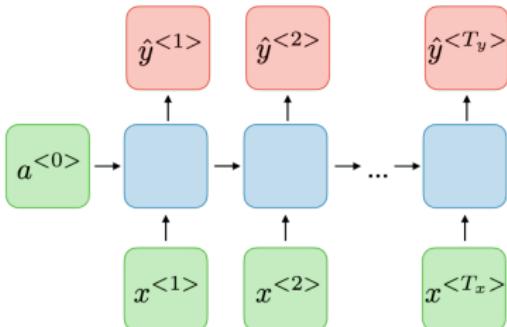


Figure: Illustration d'un RNN générique.

- Les données ont un ordre d'apparition avec une même structure.
- On garde en "mémoire" les précédentes données.

Autres types de données: brève introduction

Séquences temporelles et réseaux récurrents

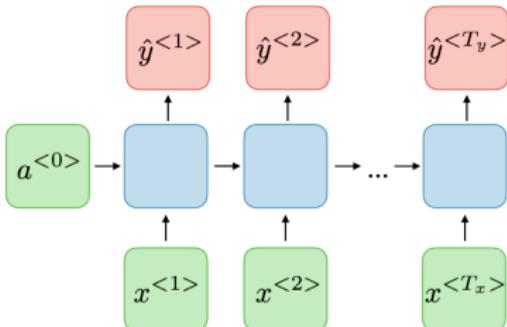


Figure: Illustration d'un RNN générique.

- Les données ont un ordre d'apparition avec une même structure.
- On garde en "mémoire" les précédentes données.
- $a_t = \tanh(W^T \text{concat}(x_t, a_{t-1}) + b)$

<https://stanford.edu/~shervine/l/fr/teaching/cs-230/pense-bete-reseaux-neurones-recurrents>

Autres types de données: brève introduction

Réseaux récurrents

Objectif: Améliorer la mémoire long-terme avec des "portes".

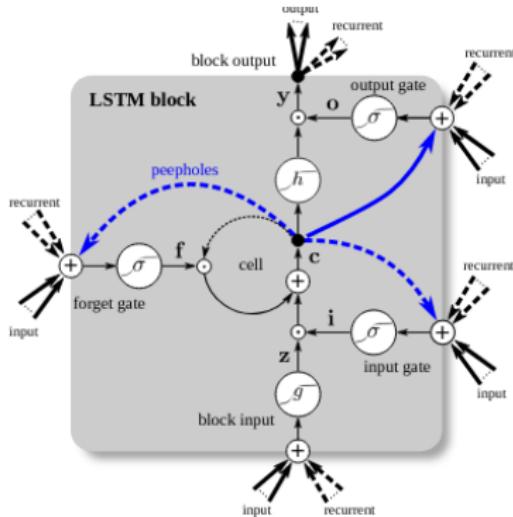


Figure: Illustration d'un LSTM.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., Schmidhuber, J. (2016). LSTM: A search space odyssey.

Autres types de données: brève introduction

Données hiérarchiques

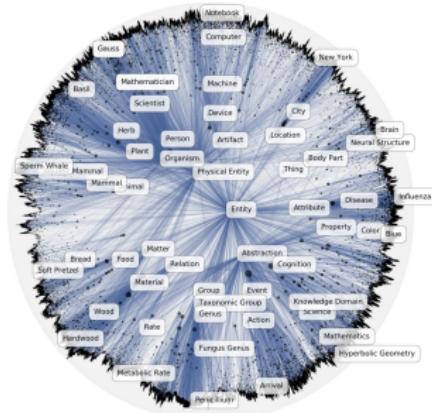


Figure: Hyperbolic embedding.

- hiérarchie dans les données: ici les relations entre mots.

Autres types de données: brève introduction

Données hiérarchiques

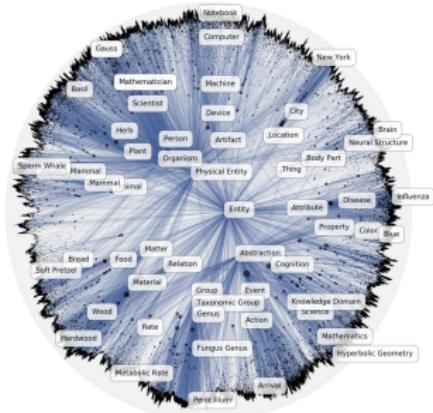


Figure: Hyperbolic embedding.

- hiérarchie dans les données: ici les relations entre mots.
 - Utilisation d'une topologie non euclidienne hyperbolique.

Autres types de données: brève introduction

Données hiérarchiques

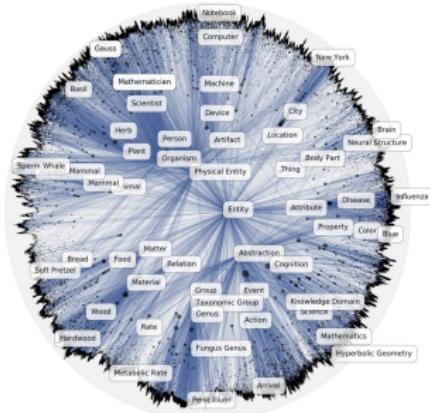


Figure: Hyperbolic embedding.

- hiérarchie dans les données: ici les relations entre mots.
 - Utilisation d'une topologie non euclidienne hyperbolique.
 - Généralisable à n dimensions (dans une hypersphère).

Nickel, M., Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations.

Autres types de données: brève introduction

Images hiérarchiques ?



Figure: Hyperbolic embedding de MNIST.

- Les symboles les plus ambigus sont vers 0 (racine).

Autres types de données: brève introduction

Graphes de données

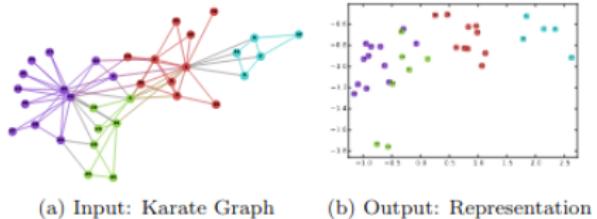


Figure: Graph embedding learned by DeepWalk (unsupervised).

- Un graphe en entrée.

Perozzi, B., Al-Rfou, R., Skiena, S. (2014, August). Deepwalk: Online learning of social representations.

Autres types de données: brève introduction

Graphes de données

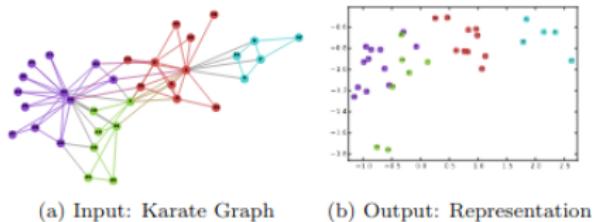


Figure: Graph embedding learned by DeepWalk (unsupervised).

- Un graphe en entrée.
- Apprend une représentation des noeuds.

Perozzi, B., Al-Rfou, R., Skiena, S. (2014, August). Deepwalk: Online learning of social representations.

Autres types de données: brève introduction

Graphes de données

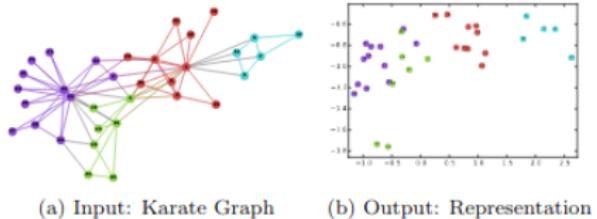


Figure: Graph embedding learned by DeepWalk (unsupervised).

- Un graphe en entrée.
- Apprend une représentation des noeuds.
- Utilise la représentation pour une tâche quelconque.

Perozzi, B., Al-Rfou, R., Skiena, S. (2014, August). Deepwalk: Online learning of social representations.

Geometric deep learning

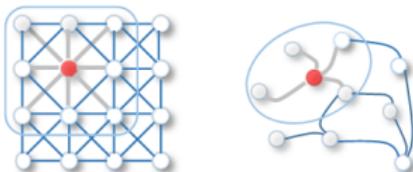


Figure: Comparaison entre une convolution 2D et une convolution de graphe.

- Soit \hat{A} la matrice d'adjacence normalisée, X^0 la représentation initiale des noeuds.

Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
 Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S. Y. (2020). A comprehensive survey on graph neural networks.

Geometric deep learning

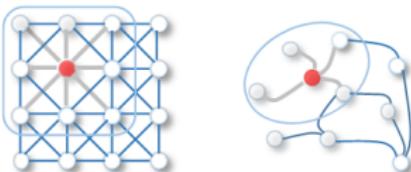


Figure: Comparaison entre une convolution 2D et une convolution de graphe.

- Soit \hat{A} la matrice d'adjacence normalisée, X^0 la représentation initiale des noeuds.
- Aggrégation des noeuds proches: $H^0 = \hat{A}X$.

Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
 Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S. Y. (2020). A comprehensive survey on graph neural networks.

Geometric deep learning

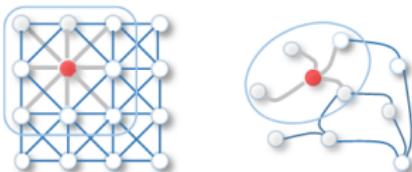


Figure: Comparaison entre une convolution 2D et une convolution de graphe.

- Soit \hat{A} la matrice d'adjacence normalisée, X^0 la représentation initiale des noeuds.
- Aggrégation des noeuds proches: $H^0 = \hat{A}X$.
- Propagation d'une couche: $X^1 = \text{ReLU}(H^0 W^0)$.

Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
 Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S. Y. (2020). A comprehensive survey on graph neural networks.

Geometric deep learning

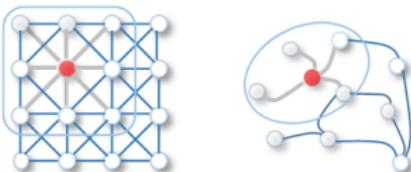


Figure: Comparaison entre une convolution 2D et une convolution de graphe.

- Soit \hat{A} la matrice d'adjacence normalisée, X^0 la représentation initiale des noeuds.
- Aggrégation des noeuds proches: $H^0 = \hat{A}X$.
- Propagation d'une couche: $X^1 = \text{ReLU}(H^0 W^0)$.
- On peut rajouter des couches.

Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
 Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S. Y. (2020). A comprehensive survey on graph neural networks.

Que faut-il retenir ?

- La structure des données détermine la méthode de *deep learning*.

Que faut-il retenir ?

- La structure des données détermine la méthode de *deep learning*.
- Toujours la back-propagation. (légère modification si on change la topologie sous-jacente).

Que faut-il retenir ?

- La structure des données détermine la méthode de *deep learning*.
- Toujours la back-propagation. (légère modification si on change la topologie sous-jacente).
- Plusieurs *hot topics*.

Que faut-il retenir ?

- La structure des données détermine la méthode de *deep learning*.
- Toujours la back-propagation. (légère modification si on change la topologie sous-jacente).
- Plusieurs *hot topics*.
- On a fait qu'effleurer la plupart des thématiques de recherche.

Autres types de données: brève introduction

Que faut-il retenir ?

- La structure des données détermine la méthode de *deep learning*.
- Toujours la back-propagation. (légère modification si on change la topologie sous-jacente).
- Plusieurs *hot topics*.
- On a fait qu'effleurer la plupart des thématiques de recherche.
- Savoir appliquer une convolution 1D et 2D.

Rappels

Table: Types d'apprentissage. Le *feedback* fait ici référence à une supervision experte.

	Avec <i>feedback</i>	Sans <i>feedback</i>
Actif	Renforcement	Motivation intrinsèque
Passif	Supervisé	Non supervisé

- Comment faire de l'apprentissage actif avec le *deep learning* ?

Rappels

Rappels - MDP

- S l'ensemble d'états;
- A l'ensemble d'actions;
- R la fonction de récompense; **inconnue**

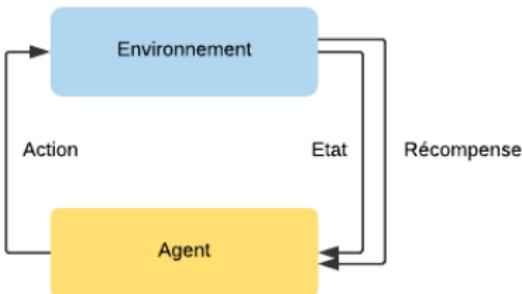


Figure: Description d'un processus de décision markovien.

Rappels

Rappels - MDP

Processus de décision markovien:

- S l'ensemble d'états;
 - A l'ensemble d'actions;
 - P la fonction de transition d'un état à l'autre; **inconnue**
 - R la fonction de récompense; **inconnue**
 - γ le facteur d'atténuation;
 - ρ_0 la distribution initiale d'état. **inconnue**

Objectif RL : Trouver la politique π^* :

$$\pi^* = \arg \max_{\pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, s_{t+1}, \pi(s_t)) \right]. \quad (1)$$

Rappels

Rappels - Équation de Bellman

On utilise l'espérance cumulée de récompense suivant un couple (état,action):

$$Q_\pi(s, a) = \mathbb{E}_{a_t \sim \pi(s_t)} \left(\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) |_{s_0=s, a_0=a} \right). \quad (2)$$

$$V_\pi(s) = \mathbb{E}_{a_t \sim \pi(s_t)} \left(\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t)) |_{s_0=s} \right). \quad (3)$$

On applique l'équation de Bellman:

$$Q_\pi(s_t, a_t) = R(s_t, a_t) + \gamma \max_a Q_\pi(P(s_t, a_t), a). \quad (4)$$

Rappels

Approche tabulaire

	s1	s2	s3	s4
a0	1	0	2	1.5
a1	0.5	0	0.1	4
a2	1	1	1	1

Idée : Table contenant les valeurs de **tous** les couples (état,action).

Rappels

Approche tabulaire

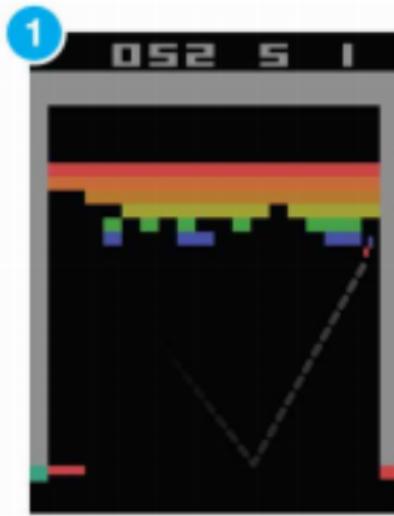
	s1	s2	s3	s4
a0	1	0	2	1.5
a1	0.5	0	0.1	4
a2	1	1	1	1

Idée : Table contenant les valeurs de **tous** les couples (état,action).

Problème : On retourne rarement dans le même état si l'espace d'état est continu ou trop grand.

Rappels

Exemple d'espace d'état trop grand



Taille de l'espace d'état:

$$|S| = 255^{84 \times 84 \times 4} = 255^{28224}$$

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$
- Choisir des caractéristiques $\phi(s, a)$ manuellement.

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$
- Choisir des caractéristiques $\phi(s, a)$ manuellement.
- Apprendre les poids θ de l'approximation linéaire.

Solution:

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$
- Choisir des caractéristiques $\phi(s, a)$ manuellement.
- Apprendre les poids θ de l'approximation linéaire.

Problèmes:

- Difficile de choisir les bonnes caractéristiques.

Solution:

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$
- Choisir des caractéristiques $\phi(s, a)$ manuellement.
- Apprendre les poids θ de l'approximation linéaire.

Problèmes:

- Difficile de choisir les bonnes caractéristiques.
- Limité par la linéarité.

Solution:

Rappels

Approximation linéaire

Idée:

- $Q(s, a) = \theta^T \phi(s, a)$
- $\theta = \theta + \alpha \times \phi(s, a) \times \delta Q$
- Choisir des caractéristiques $\phi(s, a)$ manuellement.
- Apprendre les poids θ de l'approximation linéaire.

Problèmes:

- Difficile de choisir les bonnes caractéristiques.
- Limité par la linéarité.

Solution:

- Apprendre des caractéristiques non-linéaires.

Rappels

Approches

- ➊ Méthodes "Value-based" (DQN): Choix des actions en fonction des valeurs $Q(s, a)$

Rappels

Approches

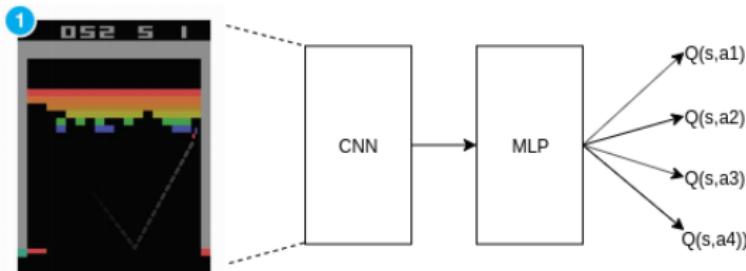
- ➊ Méthodes "Value-based" (DQN): Choix des actions en fonction des valeurs $Q(s, a)$
- ➋ Méthodes "Policy-based" (REINFORCE): Paramétrage direct de la politique $\pi_{\theta'}$.

Rappels

Approches

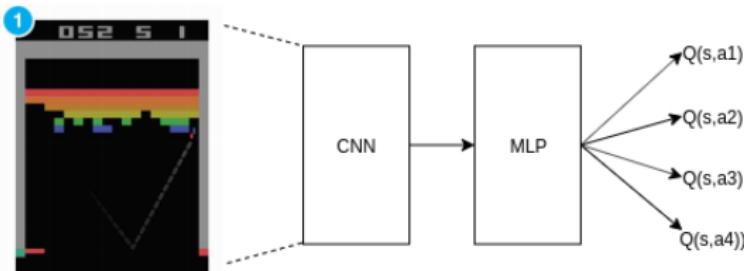
- ➊ Méthodes "Value-based" (DQN): Choix des actions en fonction des valeurs $Q(s, a)$
- ➋ Méthodes "Policy-based" (REINFORCE): Paramétrage direct de la politique $\pi_{\theta'}$.
- ➌ Méthodes "Actor-critic" (A2C): Modification de π_{θ} en fonction de $Q(s, a)$.

Deep Q-network (DQN)



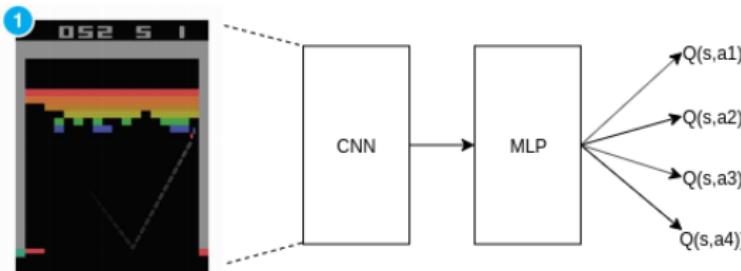
- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>

Deep Q-network (DQN)



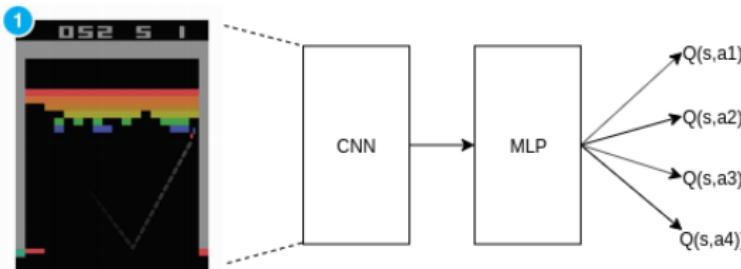
- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>
- $$G_\theta = \left[Q_\theta(s_t, a_t) - (R(s_t, a_t) + \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)) \right]^2$$

Deep Q-network (DQN)



- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>
- $G_\theta = \left[Q_\theta(s_t, a_t) - (R(s_t, a_t) + \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)) \right]^2$
- Minimisation de G_θ via backpropagation du gradient.

Deep Q-network (DQN)

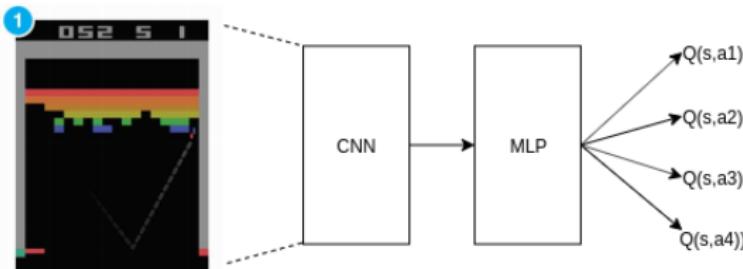


- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>
- $G_\theta = \left[Q_\theta(s_t, a_t) - (R(s_t, a_t) + \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)) \right]^2$
- Minimisation de G_θ via backpropagation du gradient.

Problèmes:

- ① \hat{Q}_θ dépend aussi de θ , pouvant faire diverger l'algorithme.

Deep Q-network (DQN)

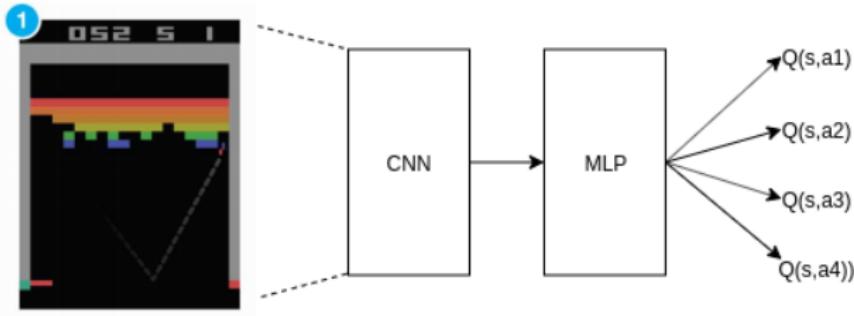


- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>
- $G_\theta = \left[Q_\theta(s_t, a_t) - (R(s_t, a_t) + \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)) \right]^2$
- Minimisation de G_θ via backpropagation du gradient.

Problèmes:

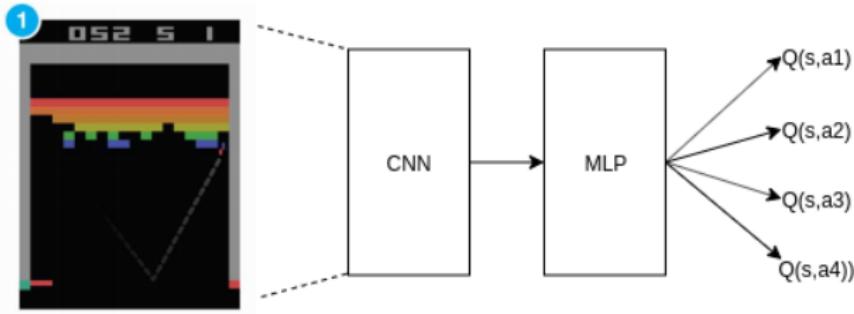
- ➊ \hat{Q}_θ dépend aussi de θ , pouvant faire diverger l'algorithme.
- ➋ Les exemples sont corrélés, rendant l'apprentissage instable.

Deep Q-network (DQN) - Astuces



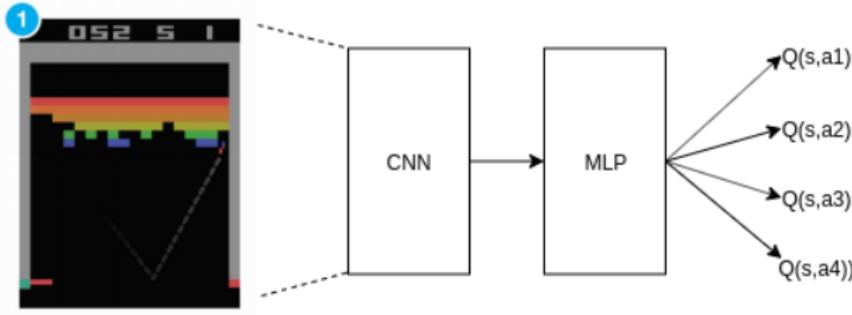
- $\|G_\theta = Q_\theta(s_t, a_t) - R(s_t, a_t) - \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)\|^2$

Deep Q-network (DQN) - Astuces



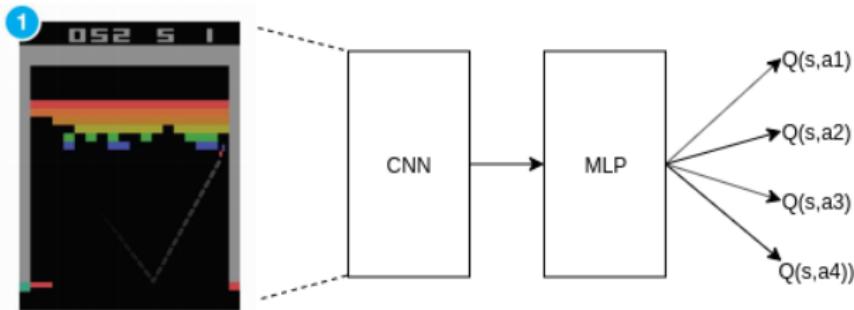
- $\|G_\theta = Q_\theta(s_t, a_t) - R(s_t, a_t) - \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)\|^2$
- Utilisation de l'experience replay : stockages des 100 000 (par exemple) dernières interactions.

Deep Q-network (DQN) - Astuces

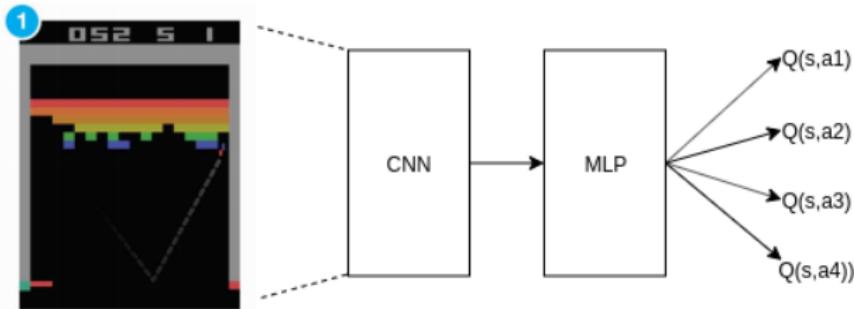


- $\|G_\theta = Q_\theta(s_t, a_t) - R(s_t, a_t) - \gamma \max_a \hat{Q}_\theta(P(s_t, a_t), a)\|^2$
- Utilisation de l'experience replay : stockages des 100 000 (par exemple) dernières interactions.
- Utilisation du target network: $\hat{Q}_{\theta'}$ est modifié graduellement vers Q_θ .

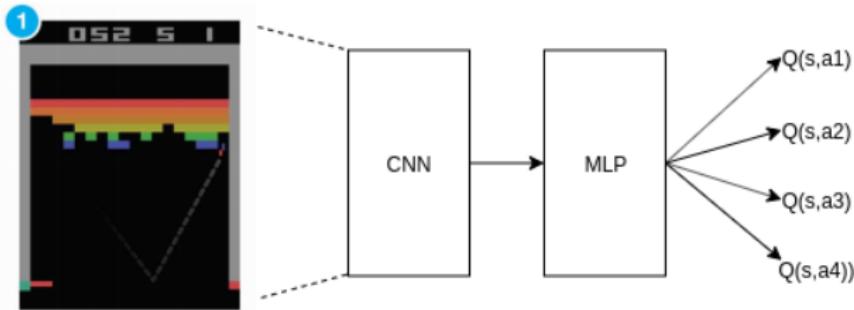
Deep Q-network (DQN) - Problème



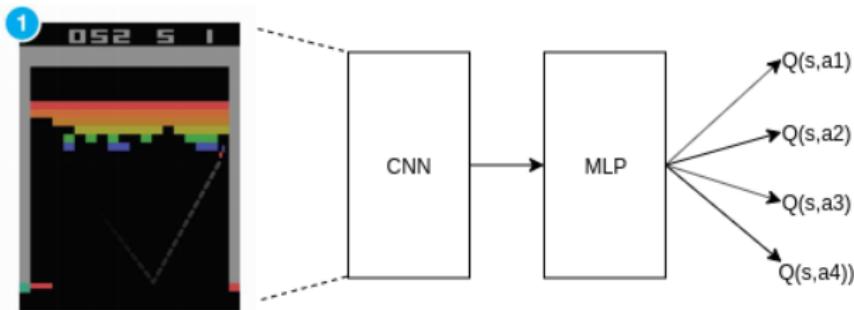
Deep Q-network (DQN) - Problème



Deep Q-network (DQN) - Problème

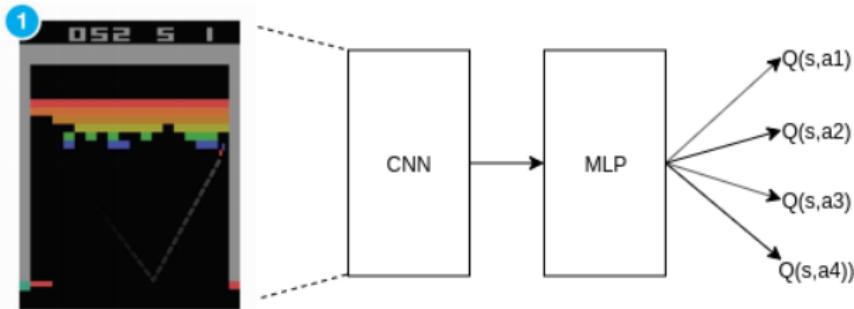


Deep Q-network (DQN) - Problème



- Le réseau calcule $Q(s, a)$ pour chaque action.

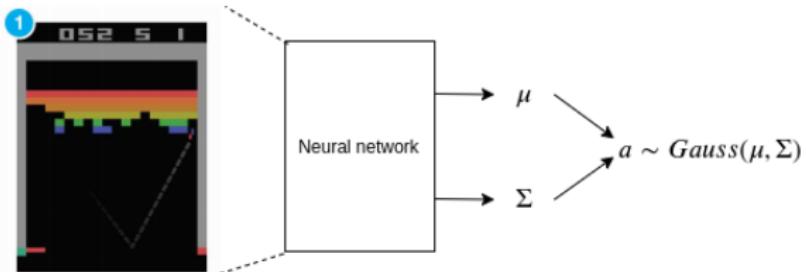
Deep Q-network (DQN) - Problème



- Le réseau calcule $Q(s, a)$ pour chaque action.
- Mais comment faire si les actions sont continues ?

Policy-based methods

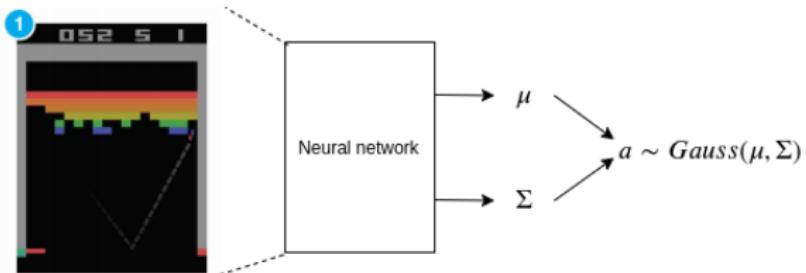
Méthodes policy-based



- Policy π génère une probabilité d'action.

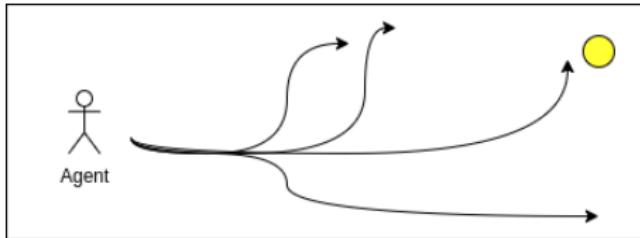
Policy-based methods

Méthodes policy-based



- Policy π génère une probabilité d'action.
- On augmente la probabilité des actions générant de fortes récompenses.

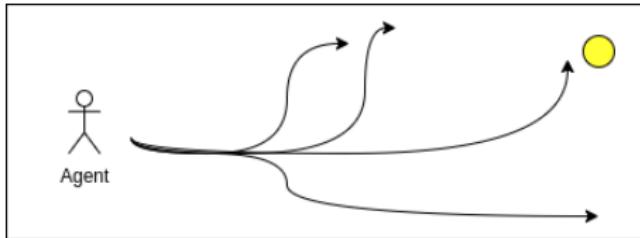
Policy-based methods

Policy gradient théorème

Objectif: π_θ maximise $J(\theta) = \mathbb{E}_{\pi_\theta} R(\tau)$.

Policy-based methods

Policy gradient théorème



Objectif: π_θ maximise $J(\theta) = \mathbb{E}_{\pi_\theta} R(\tau)$.

Trouvons le gradient: $\nabla_\theta J(\theta)$.

$$\nabla_{\theta} J(\theta) = \nabla \mathbb{E}_{\pi_{\theta}} R(\tau)$$

(5)

Policy-based methods

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\ &= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau\end{aligned}$$

(5)

Policy-based methods

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\&= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\&= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau\end{aligned}$$

(5)

Policy-based methods

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\&= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\&= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau \\&= \int \pi_{\theta}(\tau) \nabla \log \pi_{\theta}(\tau) R(\tau) d\tau.\end{aligned}\tag{5}$$

Policy-based methods

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\
 &= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \pi_{\theta}(\tau) \nabla \log \pi_{\theta}(\tau) R(\tau) d\tau. \tag{5}
 \end{aligned}$$

$$\pi_{\theta}(\tau) = \rho(s_0) \prod_1^T \pi_{\theta}(a_t | s_t) p(s_{t+1}, r_{t+1} | s_t, a_t).$$

Policy-based methods

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\
 &= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \pi_{\theta}(\tau) \nabla \log \pi_{\theta}(\tau) R(\tau) d\tau. \\
 \end{aligned} \tag{5}$$

$$\pi_{\theta}(\tau) = \rho(s_0) \prod_1^T \pi_{\theta}(a_t | s_t) p(s_{t+1}, r_{t+1} | s_t, a_t).$$

$$\log \pi_{\theta}(\tau) = C_{\theta} + \sum_1^T \log \pi_{\theta}(a_t | s_t).$$

Policy-based methods

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\
 &= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \pi_{\theta}(\tau) \nabla \log \pi_{\theta}(\tau) R(\tau) d\tau. \tag{5}
 \end{aligned}$$

$$\pi_{\theta}(\tau) = \rho(s_0) \prod_1^T \pi_{\theta}(a_t | s_t) p(s_{t+1}, r_{t+1} | s_t, a_t).$$

$$\log \pi_{\theta}(\tau) = C_{\theta} + \sum_1^T \log \pi_{\theta}(a_t | s_t).$$

$$\text{D'où } \nabla \log \pi_{\theta}(\tau) = \sum_1^T \nabla \log \pi_{\theta}(a_t | s_t).$$

Policy-based methods

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \nabla \mathbb{E}_{\pi_{\theta}} R(\tau) \\
 &= \nabla \int \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \nabla \pi_{\theta}(\tau) R(\tau) d\tau \\
 &= \int \pi_{\theta}(\tau) \nabla \log \pi_{\theta}(\tau) R(\tau) d\tau. \tag{5}
 \end{aligned}$$

$$\pi_{\theta}(\tau) = \rho(s_0) \prod_1^T \pi_{\theta}(a_t | s_t) p(s_{t+1}, r_{t+1} | s_t, a_t).$$

$$\log \pi_{\theta}(\tau) = C_{\theta} + \sum_1^T \log \pi_{\theta}(a_t | s_t).$$

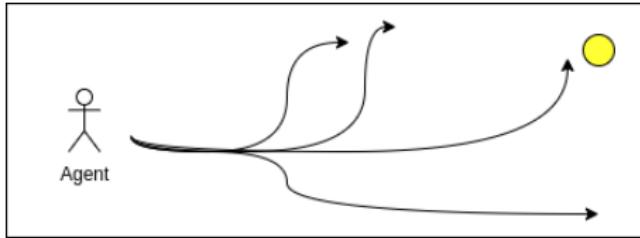
$$\text{D'où } \nabla \log \pi_{\theta}(\tau) = \sum_1^T \nabla \log \pi_{\theta}(a_t | s_t).$$

On obtient:

$$\nabla \mathbb{E}_{\pi_{\theta}} R(\tau) = \mathbb{E}_{\pi_{\theta}} R(\tau) \sum_1^T \nabla \log \pi_{\theta}(a_t | s_t).$$

Policy-based methods

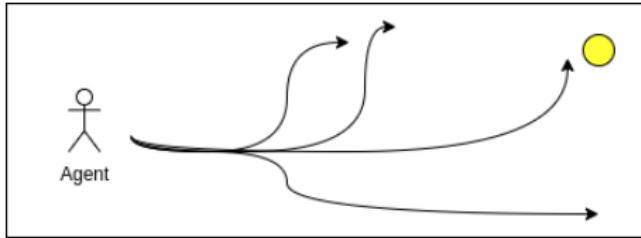
Interprétation du théorème



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$.

Policy-based methods

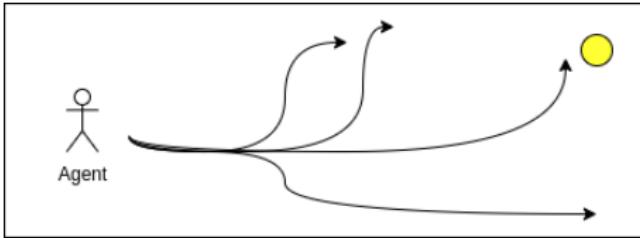
Interprétation du théorème



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$.
- Seulement besoin des trajectoires et du gradient de la politique.

Policy-based methods

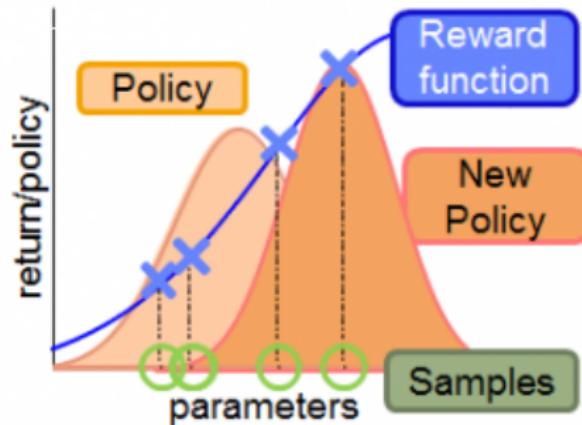
Interprétation du théorème



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$.
- Seulement besoin des trajectoires et du gradient de la politique.
- Estimateur non biaisé !

Policy-based methods

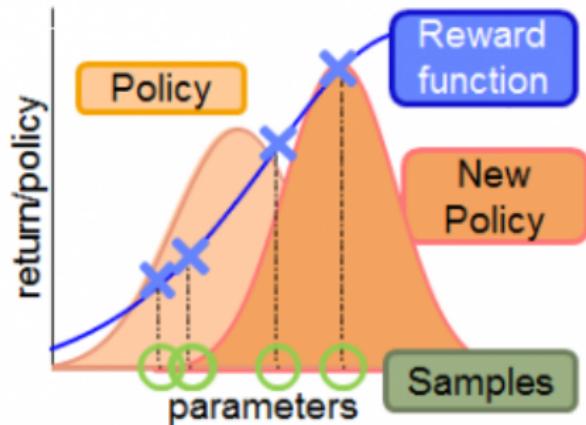
REINFORCE



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} \left[\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'}) \right].$

Policy-based methods

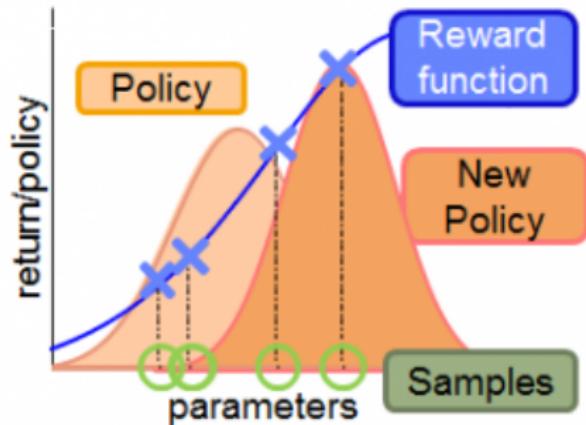
REINFORCE



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$.
- $\sum_{t'=t}^T R(s_{t'}, a_{t'})$ "pondère" la probabilité de l'action.

Policy-based methods

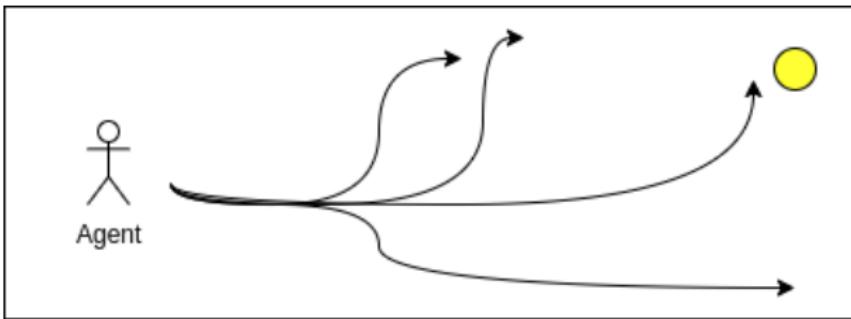
REINFORCE



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$.
- $\sum_{t'=t}^T R(s_{t'}, a_{t'})$ "pondère" la probabilité de l'action.
- Mais pour de longs épisodes, la variance est très importante.

Policy-based methods

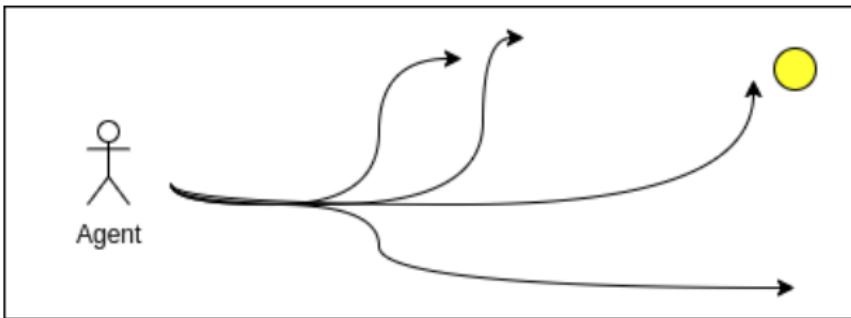
REINFORCE - Variance importante



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$
- Méthode Monte-carlo.

Policy-based methods

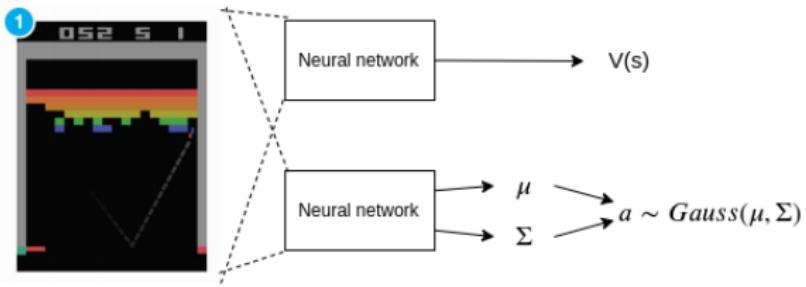
REINFORCE - Variance importante



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a, s_t \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \sum_{t'=t}^T R(s_{t'}, a_{t'})]$
- Méthode Monte-carlo.
- Variance importante.

Actor-Critic

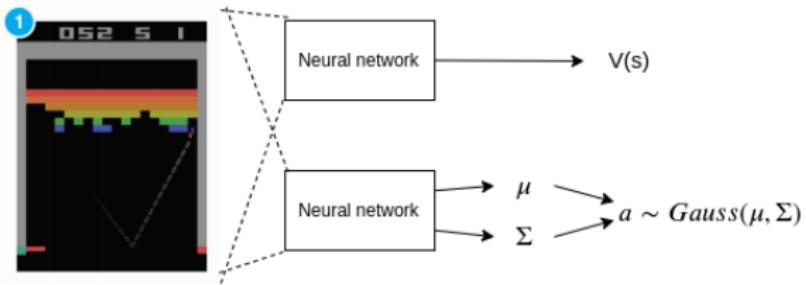
Architectures Actor-Critic



- $\sum_{t'=t}^T R(s_{t'}, a_{t'}) = R(s_t, a_t) + \gamma V_{\theta'}(s_{t+1}) = Q(s_t, a_t).$

Actor-Critic

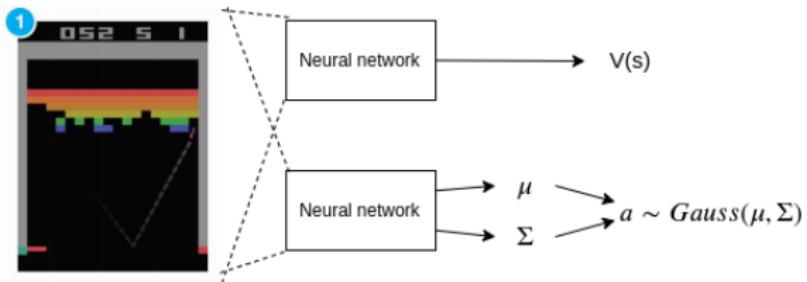
Architectures Actor-Critic



- $\sum_{t'=t}^T R(s_{t'}, a_{t'}) = R(s_t, a_t) + \gamma V_{\theta'}(s_{t+1}) = Q(s_t, a_t).$
- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) Q(s,a)].$

Actor-Critic

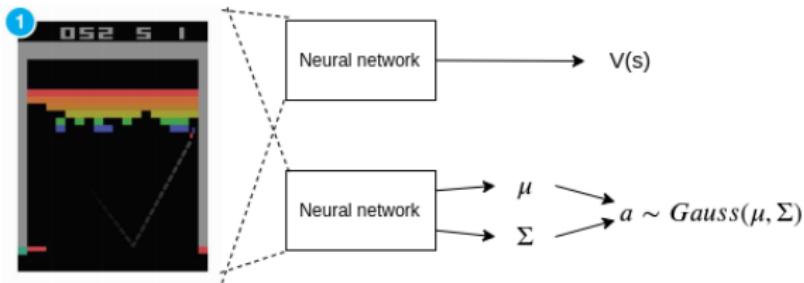
Architectures Actor-Critic



- $\sum_{t'=t}^T R(s_{t'}, a_{t'}) = R(s_t, a_t) + \gamma V_{\theta'}(s_{t+1}) = Q(s_t, a_t).$
- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) Q(s,a)].$
- Actor : Calculer $\log \pi_{\theta}(a|s)$.

Actor-Critic

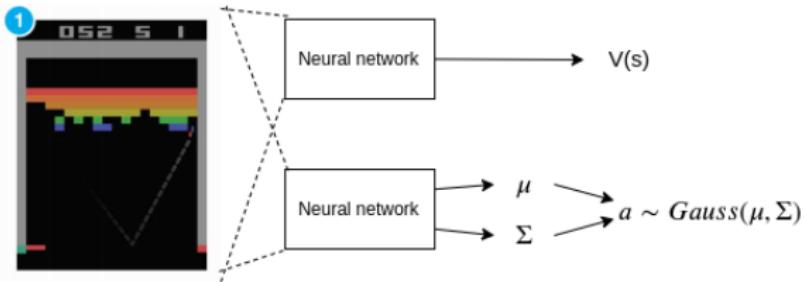
Architectures Actor-Critic



- $\sum_{t'=t}^T R(s_{t'}, a_{t'}) = R(s_t, a_t) + \gamma V_{\theta'}(s_{t+1}) = Q(s_t, a_t).$
- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) Q(s,a)].$
- Actor : Calculer $\log \pi_{\theta}(a|s)$.
- Critic : Calculer $V_{\theta'}(s')$ ou $Q_{\theta'}(s, a)$.

Actor-Critic

Réduction de la variance



$$\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s,s' \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q(s,a) - b(s))]$$

- Baseline $b(s)$ indépendante de l'action pour ne pas biaiser le gradient.

Dérivation

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q(s,a) - b(s))]$$

Actor-Critic

Dérivation

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q(s,a) - b(s))] \\ &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s)\end{aligned}$$

Actor-Critic

Dérivation

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q(s,a) - b(s))] \\ &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s) \\ &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} b(s) \nabla_{\theta} \log \pi_{\theta}(a|s)\end{aligned}$$

Actor-Critic

Dérivation

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q(s,a) - b(s))] \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} b(s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \nabla_{\theta} \log \pi_{\theta}(a|s)
 \end{aligned}$$

Actor-Critic

Dérivation

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q(s, a) - b(s))] \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} b(s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \frac{\nabla_{\theta} \pi_{\theta}(a|s)}{\pi_{\theta}(a|s)}
 \end{aligned}$$

Dérivation

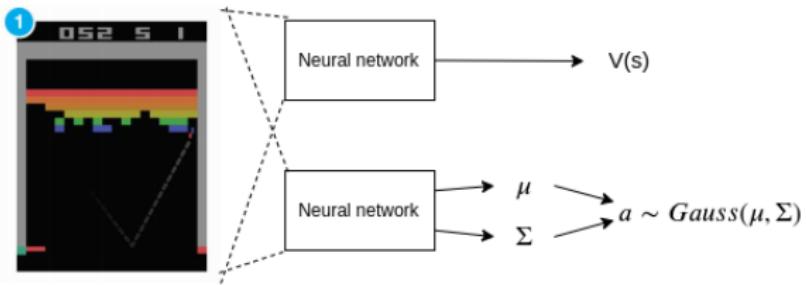
$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q(s, a) - b(s))] \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} b(s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \frac{\nabla_{\theta} \pi_{\theta}(a|s)}{\pi_{\theta}(a|s)} \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \nabla_{\theta} 1
 \end{aligned}$$

Dérivation

$$\begin{aligned}
 \nabla_{\theta} J(\theta) &= \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q(s,a) - b(s))] \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} \nabla_{\theta} \log \pi_{\theta}(a|s) b(s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \mathbb{E}_{a,s \sim \pi_{\theta}} b(s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \nabla_{\theta} \log \pi_{\theta}(a|s) \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \sum_a \pi_{\theta}(a|s) \frac{\nabla_{\theta} \pi_{\theta}(a|s)}{\pi_{\theta}(a|s)} \\
 &= \nabla_{\theta} J_{prev}(\theta) - \sum_s \mu(s) b(s) \nabla_{\theta} 1 \\
 &= \nabla_{\theta} J_{prev}(\theta)
 \end{aligned}$$

Actor-Critic

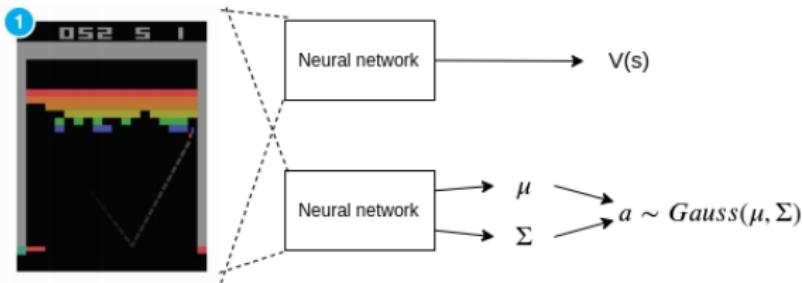
A2C: réduire la variance



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) (Q_{\theta'}(s,a) - V_{\theta'}(s))]$.

Actor-Critic

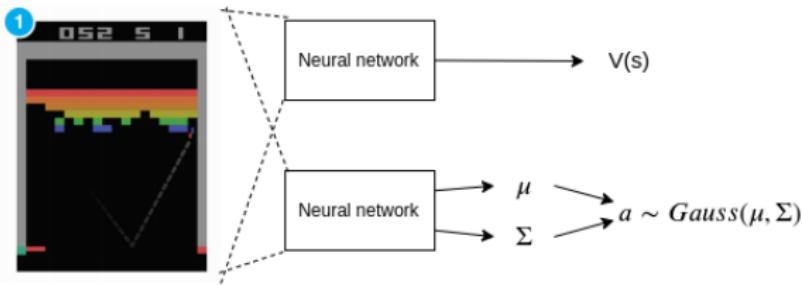
A2C: réduire la variance



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q_{\theta'}(s,a) - V_{\theta'}(s))]$.
- $A(s,a) = Q_{\theta'}(s,a) - V_{\theta'}(s)$ fonction avantage.

Actor-Critic

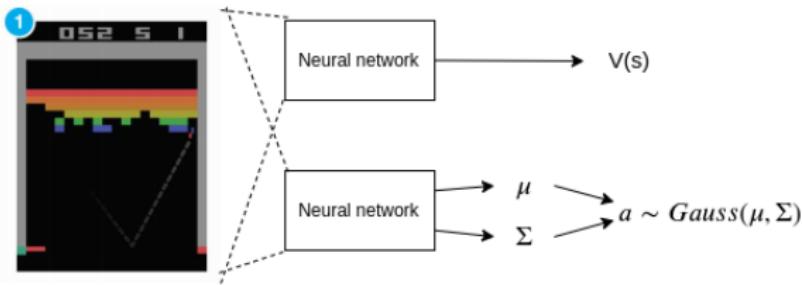
A2C: réduire la variance



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q_{\theta'}(s,a) - V_{\theta'}(s))]$.
- $A(s,a) = Q_{\theta'}(s,a) - V_{\theta'}(s)$ fonction avantage.
- $A(s,a) > 0$ si a est meilleure que la politique moyenne.

Actor-Critic

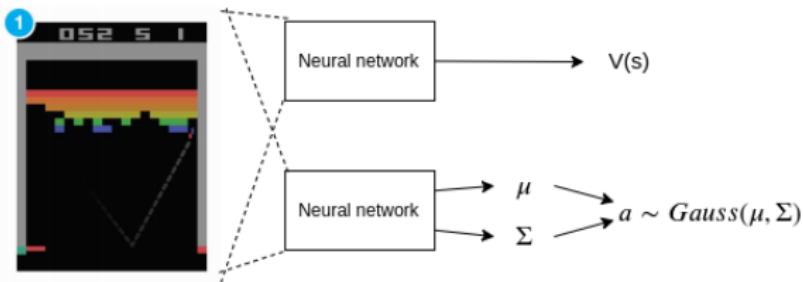
A2C: réduire la variance



- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s)(Q_{\theta'}(s,a) - V_{\theta'}(s))]$.
- $A(s,a) = Q_{\theta'}(s,a) - V_{\theta'}(s)$ fonction avantage.
- $A(s,a) > 0$ si a est meilleure que la politique moyenne.
- $A(s,a) < 0$ si a est moins bonne que la politique moyenne.

Actor-Critic

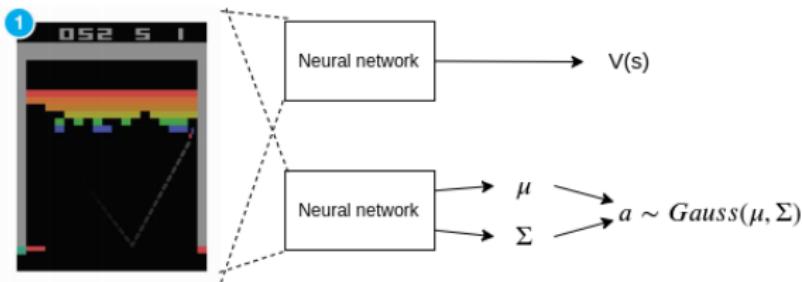
A2C: Explorer



- Convergence prématuée.

Actor-Critic

A2C: Explorer



- Convergence prématuée.
- Ajout d'un terme d'entropie forçant une forte variance.
- $\nabla_{\theta} J(\theta) = \mathbb{E}_{a,s \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) A(s, a) + \nabla_{\theta} H(\pi_{\theta}(a|s))]$.

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].
- Rainbow DQN [Hessel et al., 2018].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].
- Rainbow DQN [Hessel et al., 2018].

Modèles Actor-Critic:

- Trust region policy optimization [Schulman et al., 2015].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].
- Rainbow DQN [Hessel et al., 2018].

Modèles Actor-Critic:

- Trust region policy optimization [Schulman et al., 2015].
- Proximal policy optimization [Schulman et al., 2017].

Actor-Critic

Autres modèles...

Améliorations du DQN:

- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].
- Rainbow DQN [Hessel et al., 2018].

Modèles Actor-Critic:

- Trust region policy optimization [Schulman et al., 2015].
- Proximal policy optimization [Schulman et al., 2017].
- Deep deterministic policy gradient [Lillicrap et al., 2015].

Actor-Critic

Autres modèles...

Améliorations du DQN:

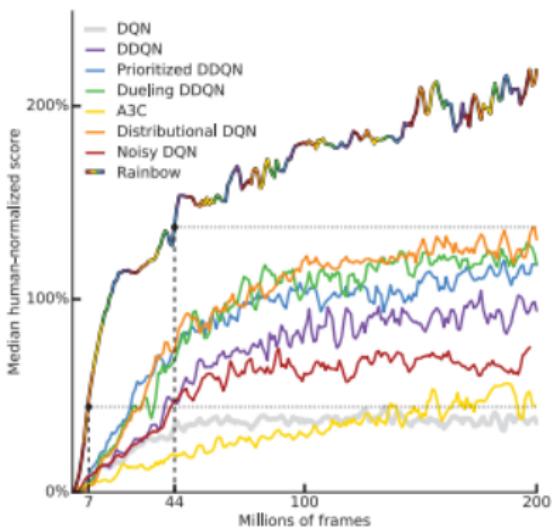
- Double DQN [Van Hasselt et al., 2016].
- Prioritized Experience replay [Schaul et al., 2015].
- Dueling DQN [Wang et al., 2015].
- Distributional DQN [Dabney et al., 2018].
- Rainbow DQN [Hessel et al., 2018].

Modèles Actor-Critic:

- Trust region policy optimization [Schulman et al., 2015].
- Proximal policy optimization [Schulman et al., 2017].
- Deep deterministic policy gradient [Lillicrap et al., 2015].
- Soft actor-critic [Haarnoja et al., 2018].

Problèmes du RL

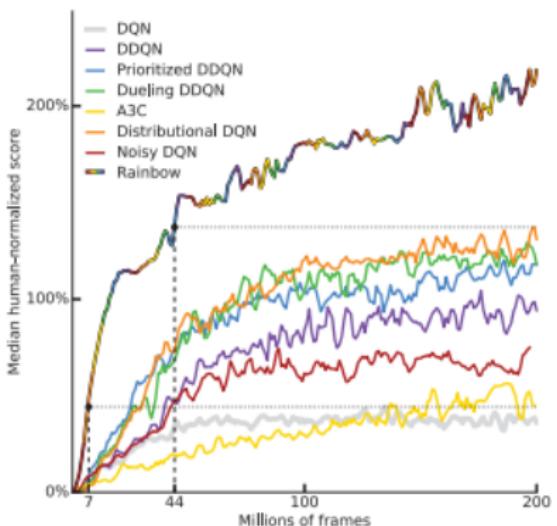
Efficacité computationnelle



- En réalité : pas de simulateurs.

Problèmes du RL

Efficacité computationnelle



- En réalité : pas de simulateurs.
- 200k frames = 1h humaine; 44M = 220h humaines.

Problèmes du RL

Exploration

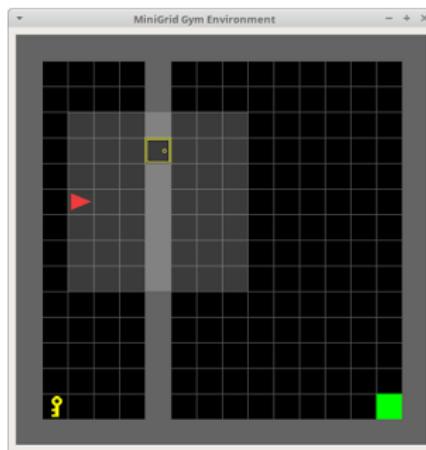
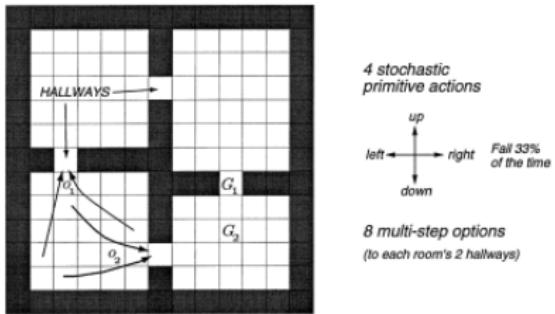


Figure: Environnement très simple avec des récompenses éparses.

- L'agent n'apprend rien car il ne trouve jamais la récompense.

Problèmes du RL

Abstraction des décisions



- Nous sommes capable de prendre des décisions de haut niveau.
- Plus facile d'apprendre sur 10 actions haut-niveau que sur 1000 actions bas-niveau [Sutton et al., 1999].

Récompenses intrinsèque

Solution

Table: Types d'apprentissage. Le *feedback* fait ici référence à une supervision experte.

	Avec <i>feedback</i>	Sans <i>feedback</i>
Actif	Renforcement	Motivation intrinsèque
Passif	Supervisé	Non supervisé

Récompenses intrinsèque

Solution

Table: Types d'apprentissage. Le *feedback* fait ici référence à une supervision experte.

	Avec <i>feedback</i>	Sans <i>feedback</i>
Actif	Renforcement	Motivation intrinsèque
Passif	Supervisé	Non supervisé

- Mixer apprentissage développemental et apprentissage par renforcement.

Récompenses intrinsèque

Solution

Table: Types d'apprentissage. Le *feedback* fait ici référence à une supervision experte.

	Avec <i>feedback</i>	Sans <i>feedback</i>
Actif	Renforcement	Motivation intrinsèque
Passif	Supervisé	Non supervisé

- Mixer apprentissage développemental et apprentissage par renforcement.
- Utiliser une récompense **intrinsèque** plutôt qu'extrinsèque.

Récompenses intrinsèque

Solution

Table: Types d'apprentissage. Le *feedback* fait ici référence à une supervision experte.

	Avec <i>feedback</i>	Sans <i>feedback</i>
Actif	Renforcement	Motivation intrinsèque
Passif	Supervisé	Non supervisé

- Mixer apprentissage développemental et apprentissage par renforcement.
- Utiliser une récompense **intrinsèque** plutôt qu'extrinsèque.
- L'agent génère lui même ses récompenses, même sans tâche.

Extrinsèque vs intrinsèque

① J'ai envie de jouer avec mes jouets !

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas.

Récompenses intrinsèque

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas.
- ⑤ Je travaille pour avoir de bonnes notes à l'école.

Récompenses intrinsèque

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas.
- ⑤ Je travaille pour avoir de bonnes notes à l'école.
- ⑥ Je veux devenir plus fort.

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas.
- ⑤ Je travaille pour avoir de bonnes notes à l'école.
- ⑥ Je veux devenir plus fort.
- ⑦ J'ai faim et je vais chercher de la nourriture.

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets !
- ② Je dois arrêter de jouer car c'est enfantin.
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas.
- ⑤ Je travaille pour avoir de bonnes notes à l'école.
- ⑥ Je veux devenir plus fort.
- ⑦ J'ai faim et je vais chercher de la nourriture.
- ⑧ J'aime découvrir et comprendre les mathématiques.

Récompenses intrinsèque

Extrinsèque vs intrinsèque

- ① J'ai envie de jouer avec mes jouets ! **Intrinsèque**
- ② Je dois arrêter de jouer car c'est enfantin. **Extrinsèque**
- ③ Je dois arrêter de jouer car les autres se moquent de moi.
Extrinsèque
- ④ Je suis excité à l'idée d'appuyer sur ce bouton magique que je ne connais pas. **Intrinsèque**
- ⑤ Je travaille pour avoir de bonnes notes à l'école.
Extrinsèque
- ⑥ Je veux devenir plus fort. **Ca dépend**
- ⑦ J'ai faim et je vais chercher de la nourriture. **Extrinsèque**
- ⑧ J'aime découvrir et comprendre les mathématiques.
Intrinsèque

Différents types de récompense intrinsèque

Curiosité

- Récompenser l'erreur de prédiction des états suivants [Pathak et al., 2017].
<https://pathak22.github.io/noreward-rl/>
- $R(s, a, s') = \|\text{forward}(e(s), a) - e(s')\|^2.$

Différents types de récompense intrinsèque

Curiosité

- Récompenser l'erreur de prédiction des états suivants [Pathak et al., 2017].
<https://pathak22.github.io/noreward-rl/>
- $R(s, a, s') = ||forward(e(s), a) - e(s')||^2$.
- Récompenser des états loins de ceux en mémoire [Savinov et al., 2018] <https://ai.googleblog.com/2018/10/curiosity-and-procrastination-in.html>.

Différents types de récompense intrinsèque

Curiosité

- Récompenser l'erreur de prédiction des états suivants [Pathak et al., 2017].
<https://pathak22.github.io/noreward-rl/>
- $R(s, a, s') = ||forward(e(s), a) - e(s')||^2.$
- Récompenser des états loins de ceux en mémoire [Savinov et al., 2018] <https://ai.googleblog.com/2018/10/curiosity-and-procrastination-in.html>.
- Récompenser l'agent selon la nouveauté des états [Bellemare et al., 2016][Ostrovski et al., 2017].
- $R(s) = \frac{1}{count(s)}.$

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ➊ Exécution d'une compétence;

<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);

<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);
- ③ Compétence récompensée selon si elle est bien classifiée;

<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);
- ③ Compétence récompensée selon si elle est bien classifiée;
- ④ Maximiser entropie des actions.

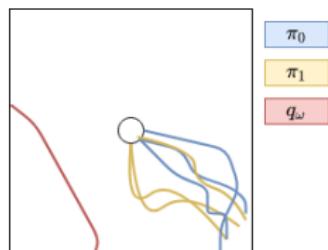
<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ➊ Exécution d'une compétence;
- ➋ Discriminateur discrimine les compétences (supervisé);
- ➌ Compétence récompensée selon si elle est bien classifiée;
- ➍ Maximiser entropie des actions.



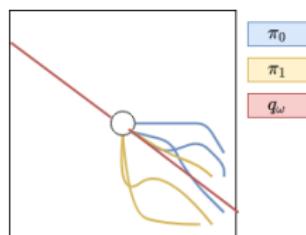
<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);
- ③ Compétence récompensée selon si elle est bien classifiée;
- ④ Maximiser entropie des actions.



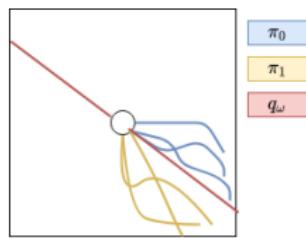
<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);
- ③ Compétence récompensée selon si elle est bien classifiée;
- ④ Maximiser entropie des actions.



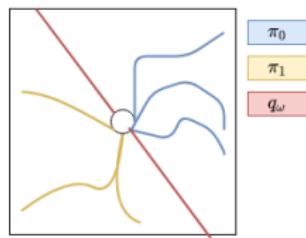
<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Génération d'objectifs

Objectif: Apprendre des compétences distinguables.

- ① Exécution d'une compétence;
- ② Discriminateur discrimine les compétences (supervisé);
- ③ Compétence récompensée selon si elle est bien classifiée;
- ④ Maximiser entropie des actions.



<https://sites.google.com/view/dads-skill/home/dads-iclr2020>

Différents types de récompense intrinsèque

Ressources et références |

Réseaux de neurones convolutionnels :

- <http://cs231n.github.io/convolutional-networks/>
- <https://towardsdatascience.com/intuitively-understanding-convolutions-for-deep-learning-1f6f42faee1>
- <https://stanford.edu/~shervine/l/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels>
- https://computersciencewiki.org/index.php/Max-pooling_-_Pooling

Apprentissage profond par renforcement:

- Excellent livre de Sutton gratuit :
<http://incompleteideas.net/book/the-book.html>
- Cours en ligne de David Silver
- Cours de Berkeley

Ressources et références II

- Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T.,
Saxton, D., and Munos, R. (2016).
Unifying count-based exploration and intrinsic motivation.
In *Advances in Neural Information Processing Systems*,
pages 1471–1479.
 - Dabney, W., Ostrovski, G., Silver, D., and Munos, R.
(2018).
Implicit quantile networks for distributional reinforcement
learning.
arXiv preprint arXiv:1806.06923.

Différents types de récompense intrinsèque

Ressources et références III

-  Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018).
Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor.
arXiv preprint arXiv:1801.01290.
-  Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., and Silver, D. (2018).
Rainbow: Combining improvements in deep reinforcement learning.
In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Différents types de récompense intrinsèque

Ressources et références IV



Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2015).
Continuous control with deep reinforcement learning.
arXiv preprint arXiv:1509.02971.



Ostrovski, G., Bellemare, M. G., Oord, A. v. d., and Munos, R. (2017).
Count-based exploration with neural density models.
arXiv preprint arXiv:1703.01310.



Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. (2017).
Curiosity-driven exploration by self-supervised prediction.
In *International Conference on Machine Learning (ICML)*, volume 2017.

Différents types de récompense intrinsèque

Ressources et références V

-  Savinov, N., Raichuk, A., Marinier, R., Vincent, D., Pollefeyns, M., Lillicrap, T., and Gelly, S. (2018).
Episodic curiosity through reachability.
arXiv preprint arXiv:1810.02274.
-  Schaul, T., Quan, J., Antonoglou, I., and Silver, D. (2015).
Prioritized experience replay.
arXiv preprint arXiv:1511.05952.
-  Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015).
Trust region policy optimization.
In *International conference on machine learning*, pages 1889–1897.

Différents types de récompense intrinsèque

Ressources et références VI

-  Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017).
Proximal policy optimization algorithms.
arXiv preprint arXiv:1707.06347.
 -  Sutton, R. S., Precup, D., and Singh, S. (1999).
Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning.
Artificial intelligence, 112(1-2):181–211.
 -  Van Hasselt, H., Guez, A., and Silver, D. (2016).
Deep reinforcement learning with double q-learning.
In *Thirtieth AAAI conference on artificial intelligence*.

Différents types de récompense intrinsèque

Ressources et références VII



Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., and De Freitas, N. (2015).

Dueling network architectures for deep reinforcement learning.

arXiv preprint arXiv:1511.06581.