

# L'apprentissage algorithmique, une nouvelle étape pour l'IA. Une application aux opérations arithmétiques

Frédéric Armetta, Anthony Baccuet, Mathieu Lefort  
Laboratoire LIRIS  
{frederic.armetta, mathieu.lefort}@univ-lyon1.fr

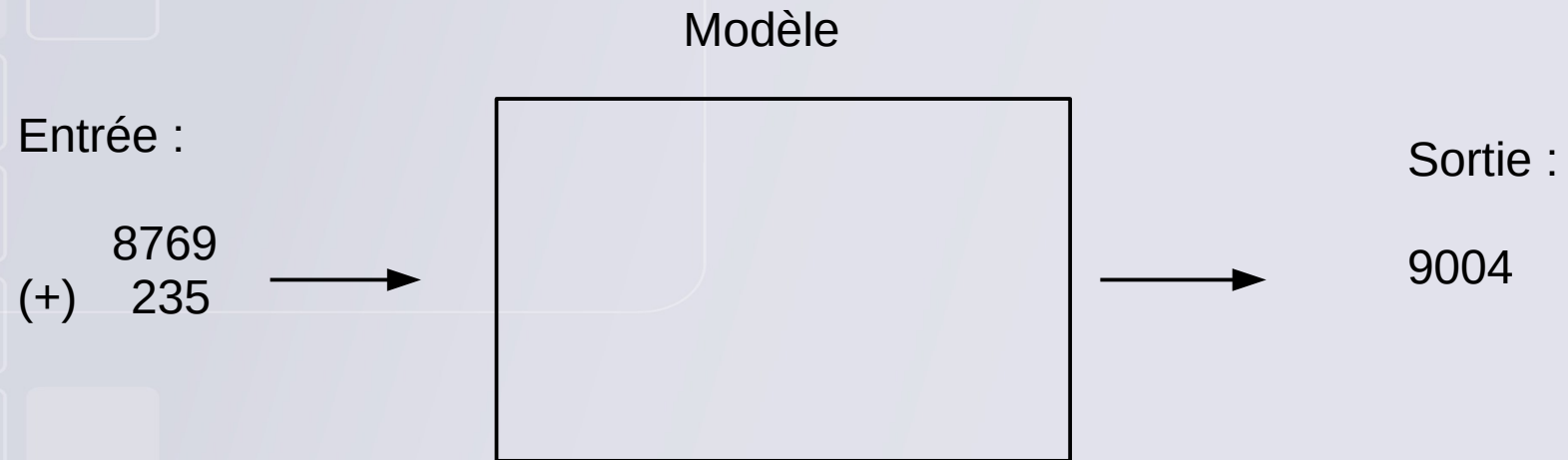
# Apprentissage d'algorithmes

- Un algorithme :
  - des **paramètres d'entrée**
  - des **variables**
  - une succession d'**opérations** de transformation menant à un **résultat**
- Problème abordé :
  - Apprentissage **de bout-en-bout** par un réseau de neurones
  - Apprentissage **supervisé**
  - Apprendre les transformations permettant de calculer à partir de paramètres d'entrée le résultat attendu

# Apprentissage d'algorithmes

- Un problème d'**apprentissage difficile**
  - Dépendances à long terme, gradient évanescent, large espace d'exploration
  - Manipulation symboliques et apprentissage statistique
- Les **opérations arithmétiques décimales**
  - Opérations de difficulté variable (addition, soustraction, multiplication, etc.)
  - Pour les plus complexes : constitue un défi, de nombreux modèles ne parviennent pas à les apprendre correctement

# Une tâche d'apprentissage simple : l'addition



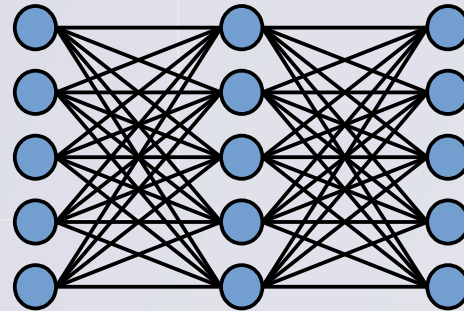
# Une tâche d'apprentissage simple : l'addition

Entrée :

8769  
(+) 235



Modèle



Sortie :

9004

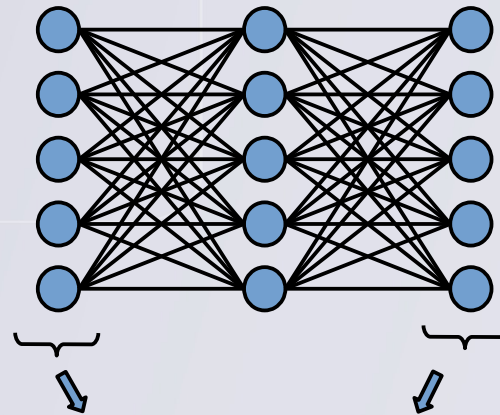
# Une tâche d'apprentissage simple : l'addition

Entrée :

8769  
(+) 235



Modèle



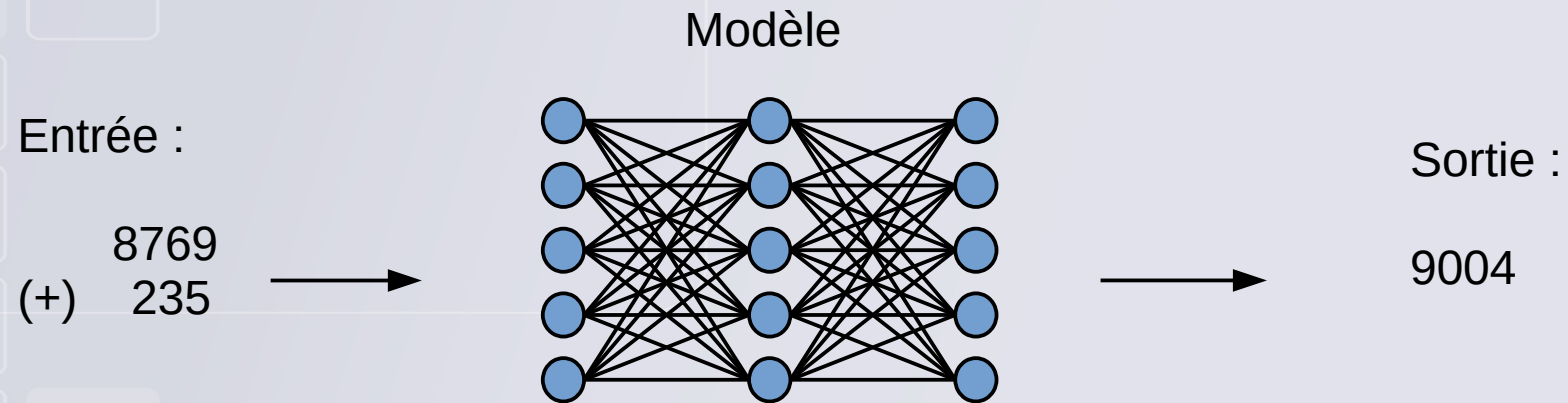
Sortie :

9004

Encodage one-hot

0	1	2	3	4	5	6	7	8	9
●	○	○	○	○	○	○	○	○	○
○	●	○	○	○	○	○	○	○	○
○	○	●	○	○	○	○	○	○	○
○	○	○	●	○	○	○	○	○	○
○	○	○	○	●	○	○	○	○	○
○	○	○	○	○	●	○	○	○	○
○	○	○	○	○	○	●	○	○	○
○	○	○	○	○	○	○	●	○	○
○	○	○	○	○	○	○	○	●	○
○	○	○	○	○	○	○	○	○	●

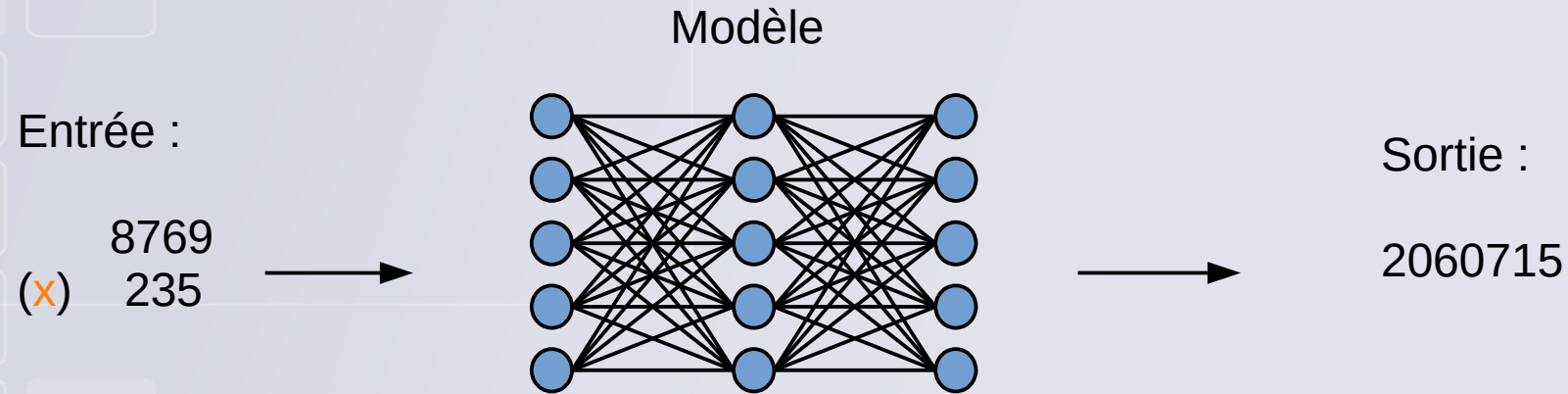
# Une tâche d'apprentissage simple : l'addition



	Opération	% Erreur
Hoshen et al.	Addition	1,7 %

[Hoshen et al.] Yedid Hoshen and Shmuel Peleg. Visual learning of arithmetic operations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, pages 3733–3739. AAAI Press, 2016.

# La multiplication décimale



	Opération	% Erreur
Hoshen et al.	Addition	1,7 %
Hoshen et al.	Multiplication	37.6%

[Hoshen et al.] Yedid Hoshen and Shmuel Peleg. Visual learning of arithmetic operations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, pages 3733–3739. AAAI Press, 2016.



# Modèle utilisé : Seq2Seq

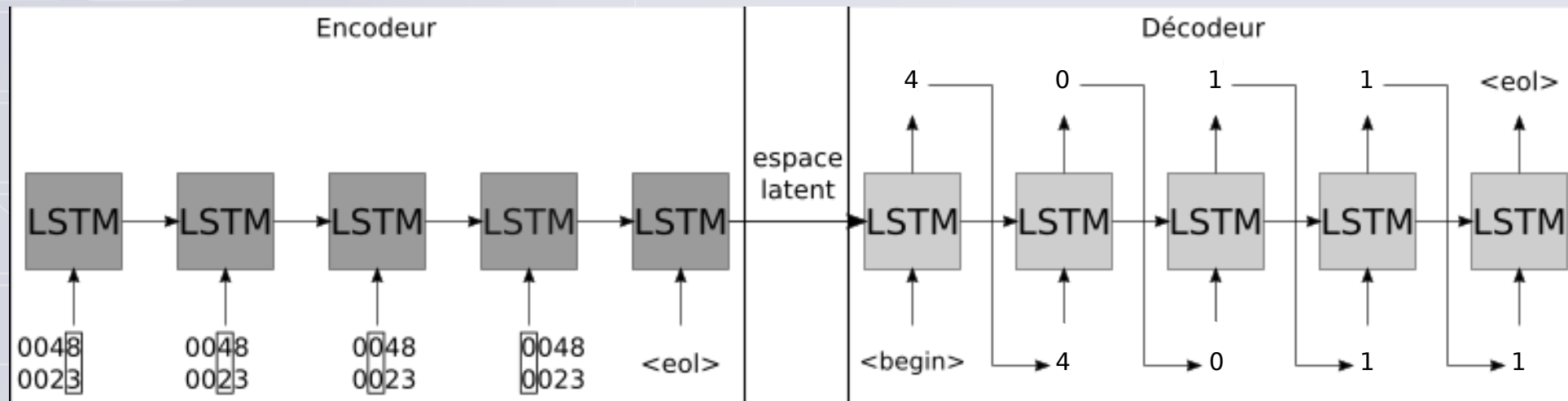


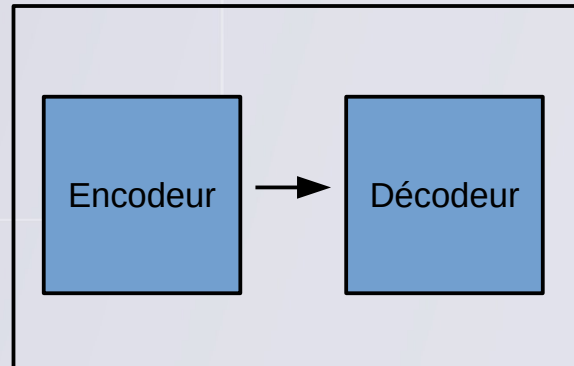
Illustration :  $48 * 23 = 1104$

# Modèle Seq2Seq

Modèle utilisé pour UAT : Unrolling Algorithmic Training  
(de type Seq2Seq)

Entrée :

(x) 8769  
235



Sortie :

2060715

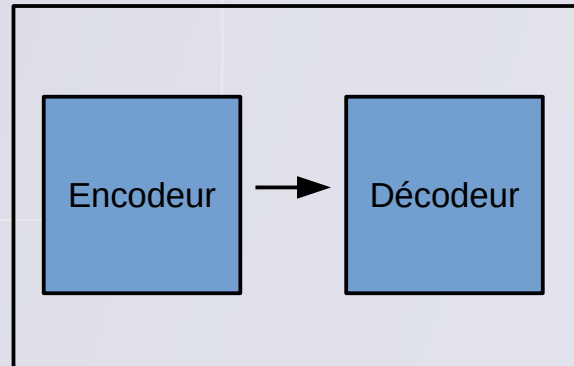
	Opération	% Erreur
Hoshen et al.	Multiplication <b>Max( Sortie )=7</b>	37.6%
Modèle UAT sans apprentissage actif	Multiplication <b>Max( Sortie )=7</b>	35.87%

# Modèle Seq2Seq

Modèle utilisé pour UAT : Unrolling Algorithmic Training  
(de type Seq2Seq)

Entrée :

(x) 8769  
235



Sortie :

2060715

	Opération	% Erreur
Hoshen et al.	Multiplication <b>Max( Sortie )=7</b>	37.6%
Modèle UAT sans apprentissage actif	Multiplication <b>Max( Sortie )=7</b>	35.87%
Modèle UAT sans apprentissage actif	Multiplication <b>Max( Sortie )=8</b>	92.42%

# UAT : Unrolling Algorithmic Training

$$23 \times 48 = 1104$$

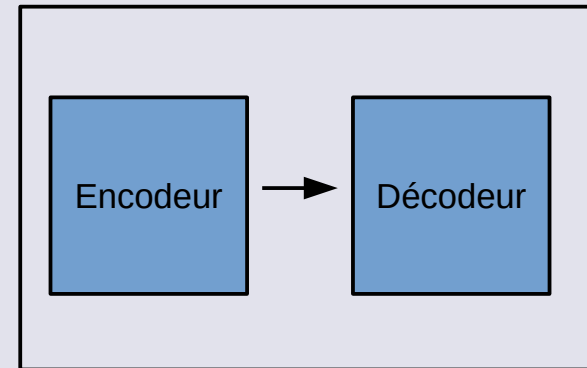
	00 <b>23</b>	(1)
×	00 <b>48</b>	(2)
<hr/>		
	0012	(3)
	0184	(4)
	0010	(5)
+	0920	(6)
<hr/>		
	0110	(7)
	<b>1104</b>	(8)

# Tâches d'apprentissage

$$23 \times 48 = 1104$$

×	0023	(1)		<b>E</b>
	0048	(2)		
<hr/>				
	0012	(3)		<b>D</b>
	0184	(4)		
	0010	(5)		
	0920	(6)		
+	<hr/>			
	0110	(7)		
	1104	(8)		

st1 (1//2)  
st2 (1//2) (3//4)  
st3 (1//2) (3//4) (5//6)  
tglobale (1//2) (lignes vides)



(3//4)  
(5//6)  
(7//8)  
(7//8)

# Tâches d'apprentissage

$$23 \times 48 = 1104$$

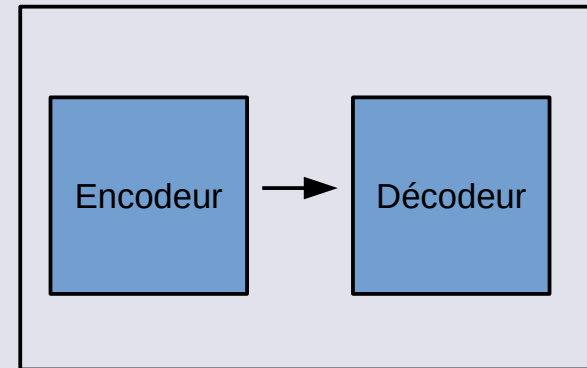
×	0023	(1)	
	0048	(2)	<b>E</b>
	0012	(3)	
	0184	(4)	
+	0010	(5)	
	0920	(6)	<b>D</b>
	0110	(7)	
	1104	(8)	

st1 (1//2)

**st2 (1//2) (3//4)**

st3 (1//2) (3//4) (5//6)

tglobale (1//2) (lignes vides)



(3//4)

**(5//6)**

(7//8)

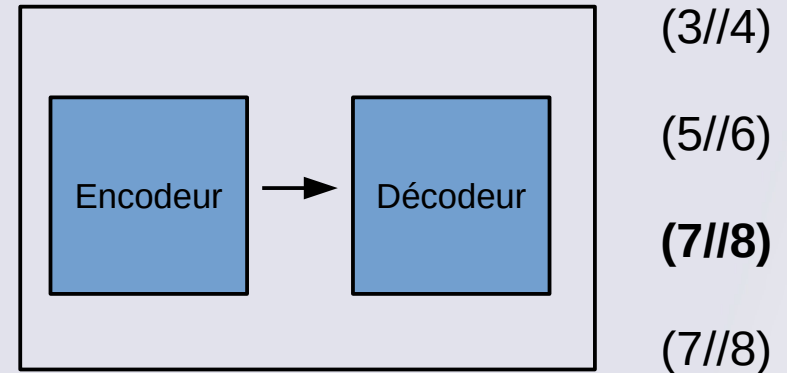
(7//8)

# Tâches d'apprentissage

$$23 \times 48 = 1104$$

	0023	(1)	
×	0048	(2)	<b>E</b>
<hr/>			
	0012	(3)	
	0184	(4)	
	0010	(5)	
+	0920	(6)	
<hr/>			
	0110	(7)	
	1104	(8)	<b>D</b>

st1 (1//2)  
st2 (1//2) (3//4)  
**st3 (1//2) (3//4) (5//6)**  
tglobale (1//2) (lignes vides)



# Tâches d'apprentissage

$$23 \times 48 = 1104$$

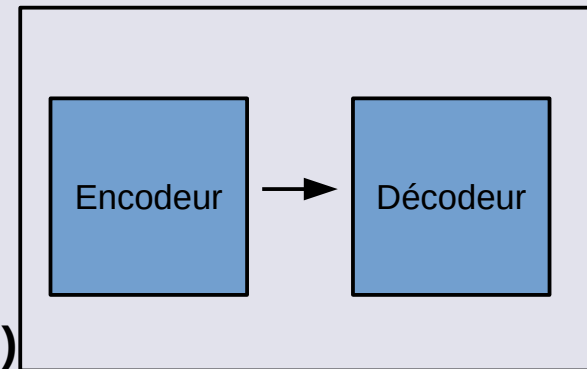
×	0023	(1)	<b>E</b>
	0048	(2)	
	« »		
	« »		
	« »		
+	« »		
	0110	(7)	<b>D</b>
	1104	(8)	

st1 (1//2)

st2 (1//2) (3//4)

st3 (1//2) (3//4) (5//6)

**tglobale (1//2) (lignes vides)**



(3//4)

(5//6)

(7//8)

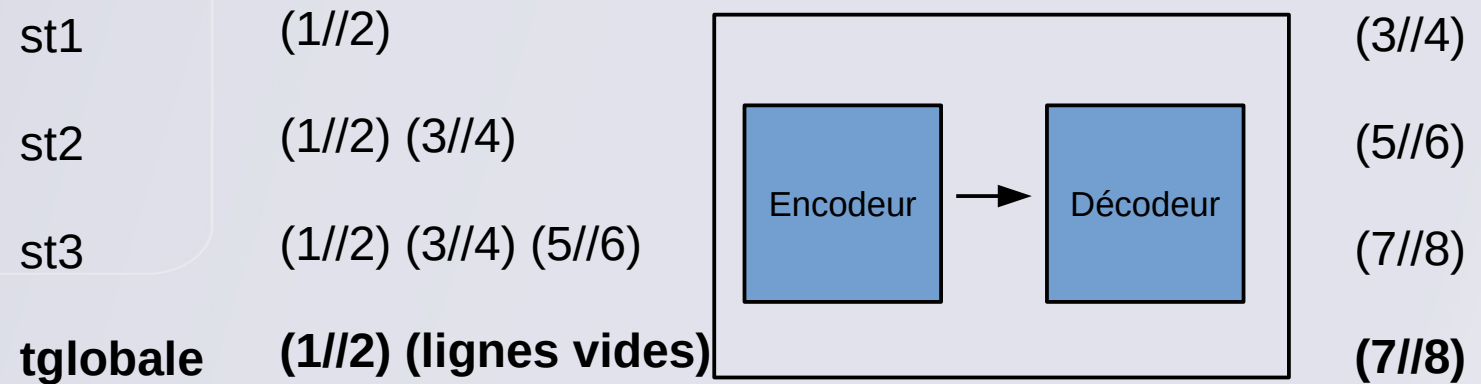
**(7//8)**



# Apprentissage actif

23 x 48 = 1104

×	0023 (1)	E
	0048 (2)	
	« »	
	« »	
	« »	
+	« »	
	0110 (7)	D
	1104 (8)	

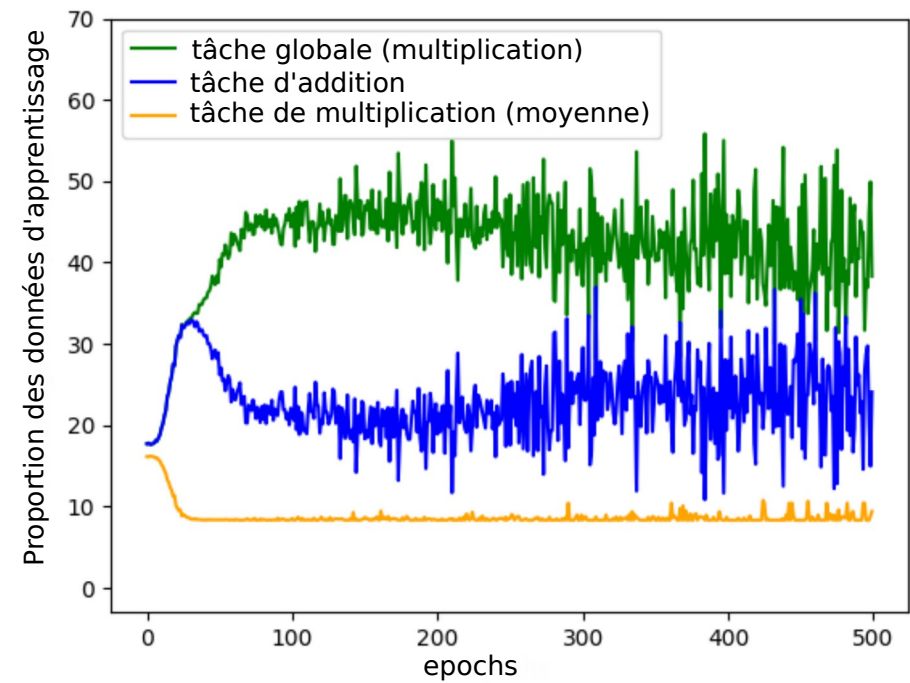
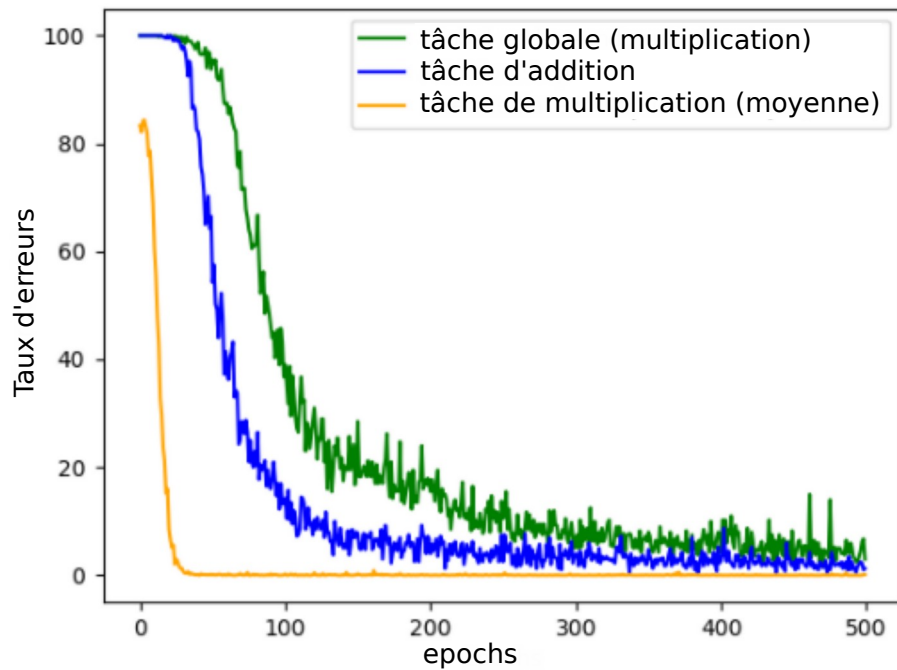


$$F_{t\grave{a}che} = \lambda \frac{err_{t\grave{a}che}}{\sum_{ta \in listeT\grave{a}che} err_{ta}} + (1 - \lambda) \frac{1}{card(listeT\grave{a}che)}$$

[Nollet et al.]

[Nollet et al.] Bastien Nollet, Mathieu Lefort, and Frédéric Armetta. Learning Arithmetic Operations With A Multistep Deep Learning. In The International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, July 2020.

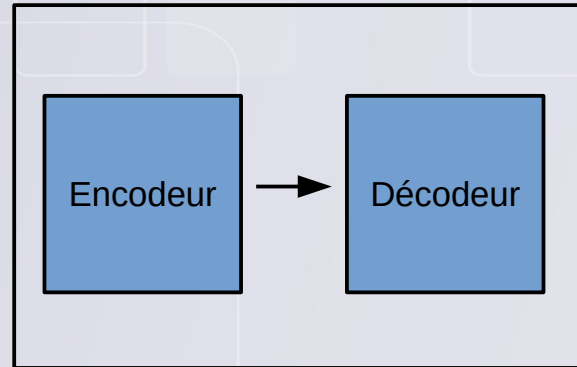
# Apprentissage actif



# Influence de la méthode d'apprentissage

Entrée :

8769  
(x) 235



Sortie :

2060715

	Opération	% Erreur (tglobale)
UAT sans → avec apprentissage actif	Multiplication Max( Sortie )=7	35.87% → 4.51% (± 1.21%)
UAT sans → avec apprentissage actif	Multiplication Max( Sortie )=8	92.42% → 23.34% (± 14.69%)

## Fine tuning sur tglobal

Multiplication Max( Sortie )=8	% Erreur (tglobale)
UAT avec apprentissage actif	(92.42% →) 23.34% (± 14.69%)
initial UAT train = sous-tâches + tglobale	(23.34% →) <b>6,68 %</b> (± <b>4.31%</b> )
initial UAT train = sous-tâches	73.31% (± 11.10%)

→ une fois amorcée, tglobal peut être apprise seule

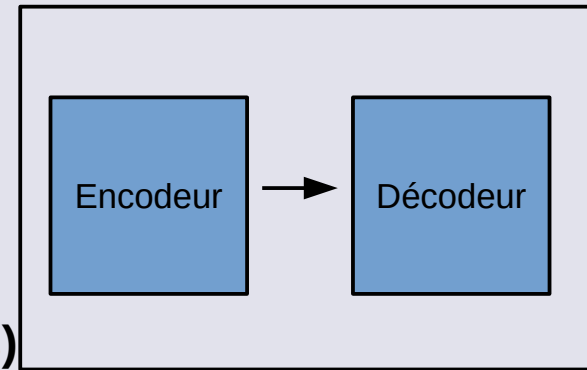
→ tglobal est nécessaire pour le *pre-training*

# Faire varier les récurrences disponibles (*fine tuning*)

$$23 \times 48 = 1104$$

×	0023	(1)	<b>E</b>
	0048	(2)	
	« »		
	« »		
	« »		
+	« »		
	0110	(7)	<b>D</b>
	1104	(8)	

st1 (1//2)  
 st2 (1//2) (3//4)  
 st3 (1//2) (3//4) (5//6)  
 tglobale (1//2) (**lignes vides**)



(3//4)  
 (5//6)  
 (7//8)  
 (7//8)

in [0, 7]

# Fine tuning et variation de la quantité de lignes vides encodées

Tglobale : (1//2) (lignes vides) → Sortie

Lignes doubles vides	% Erreur (tglobale)
0	88,67 % (± 5.03%)
1	50,19 % (± 18.16%)
2	14,39 % (± 2.65%)
3	4,44 % (± 1.65%)
4	6,68 % (± 4.31%)
5	5,53 % (± 2.86%)
6	3,84 % (± 1.61%)
7	<b>3,83 % (± 1.29%)</b>

Réurrences disponibles



# Conclusion

- Des difficultés d'entraînement propres à l'apprentissage d'algorithmes, tels que la multiplication (complexité de la tâche, dépendances à long terme)
- UAT (Unrolling Algorithmic Training) : une méthode originale qui guide l'apprentissage par l'introduction de sous-tâches algorithmiques
  - L'amélioration des performances mesurée est directement causée par la méthode d'apprentissage
  - Transfert algorithmique
  - Amorçe de la tâche d'apprentissage globale puis *fine tuning*
  - Robustesse et adaptation du modèle lorsque les récurrences disponibles varient

# Perspectives

- Étude du processus de transfert. Curriculum learning ?
- Une généralisation possible a des algorithmes proches ?
- Quelles applications ? Les modèles de langage ?