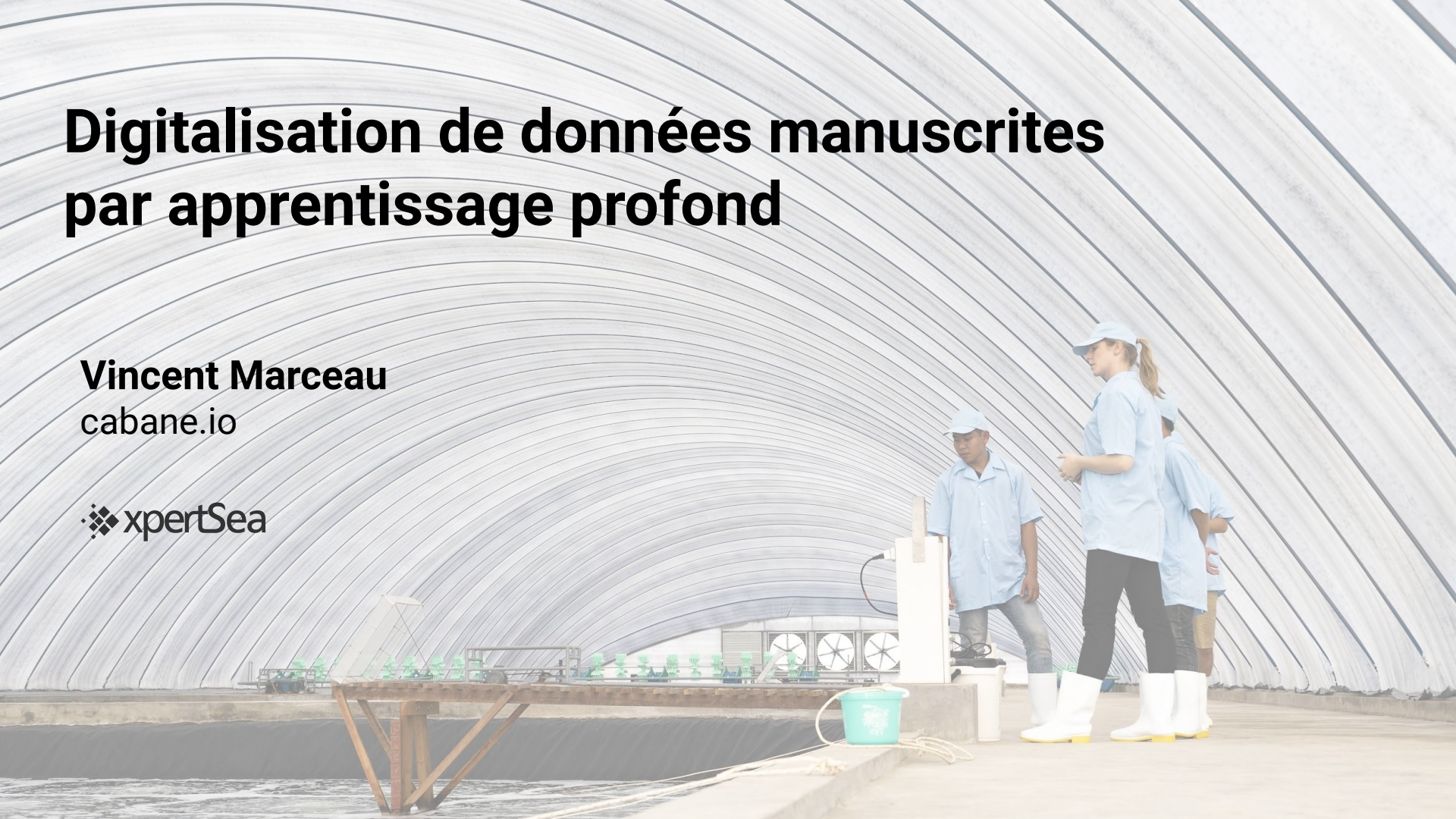
Digitalisation de données manuscrites par apprentissage profond

Vincent Marceau
cabane.io

xpertSea

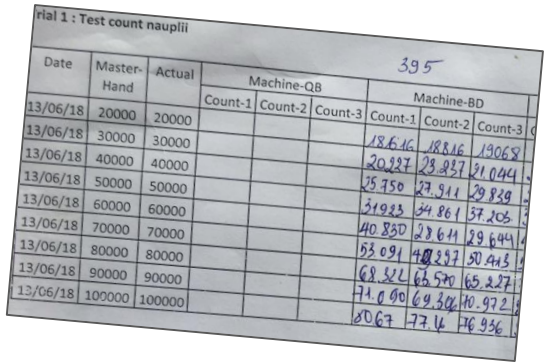


Digitalisation de données

Pourquoi?
Faciliter l’entrée de données sur notre plateforme.

Comment?
Digitaliser des logs manuscrits des fermiers.

Quoi?
Développer un modèle de reconnaissance de texte manuscrit.



Trial 1 : Test count nauplii

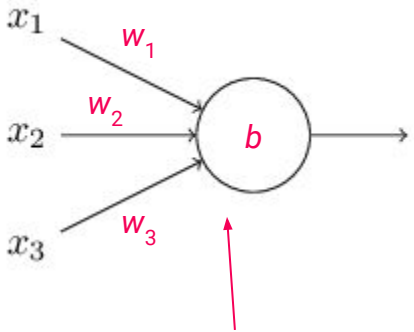
395

Date	Master-Hand	Actual	Machine-QB			Machine-BD		
			Count-1	Count-2	Count-3	Count-1	Count-2	Count-3
13/06/18	20000	20000				18.616	18.816	19068
13/06/18	30000	30000				20.227	28.837	31.044
13/06/18	40000	40000				25.750	27.911	29.839
13/06/18	50000	50000				31.923	34.861	37.205
13/06/18	60000	60000				40.850	28.611	29.644
13/06/18	70000	70000				53.091	48.337	50.443
13/06/18	80000	80000				68.321	66.570	65.227
13/06/18	90000	90000				71.090	69.306	70.972
13/06/18	100000	100000				80.67	77.4	76.936

Réseaux de neurones: récapitulatif

Le neurone sigmoide

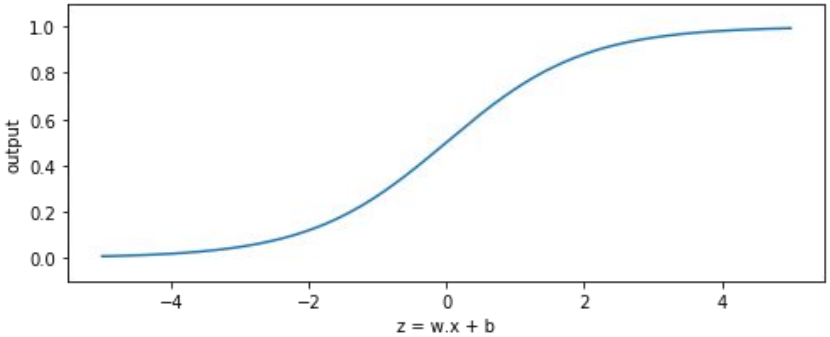
inputs



fonction d'activation

$$output = \sigma(z) = 1/(1+e^{-z}) \qquad z \equiv \mathbf{w}\cdot\mathbf{x} + b = \sum w_j x_j + b$$

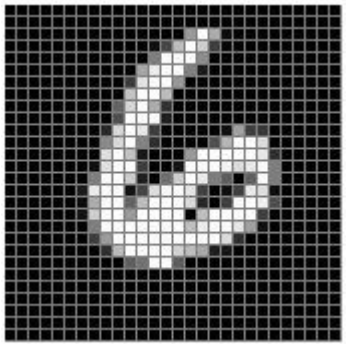
Paramètres du neurone

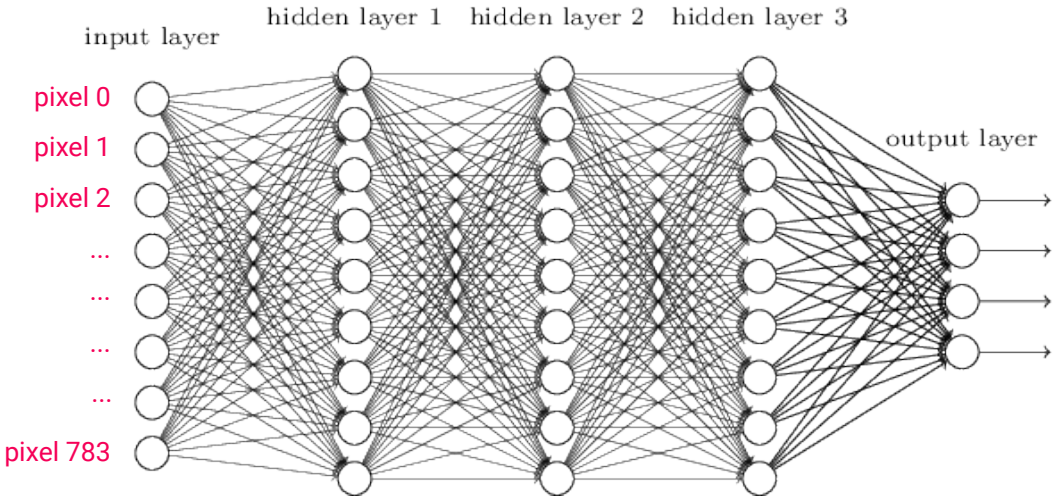


Réseaux de neurones: récapitulatif

Réseau de neurones “fully connected”

Image d’entrée (MNIST)





Distribution de probabilité des caractères de 0 à 9

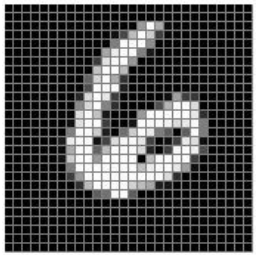
Image: M. Nielsen, Neural networks and deep learning

Réseaux de neurones: récapitulatif

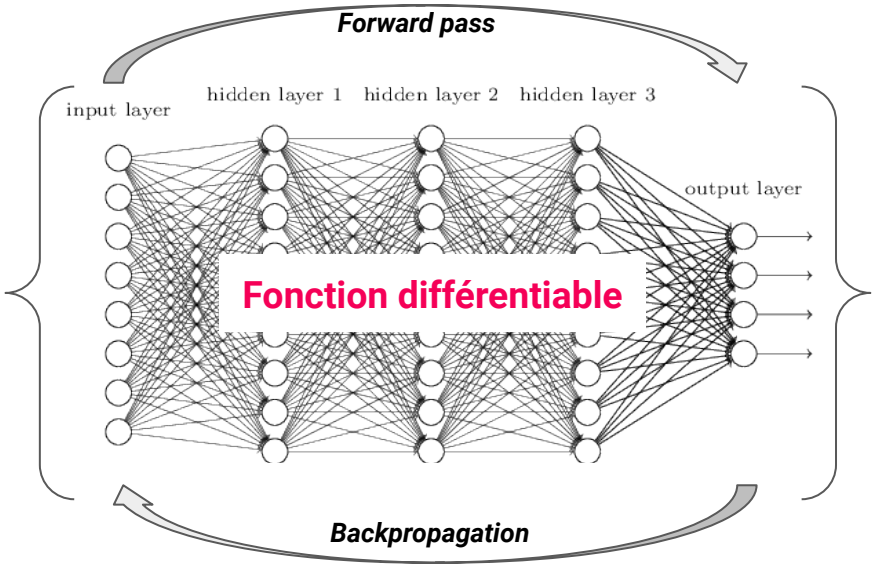
Entraînement par descente du gradient

Données d'entraînement

Image annotée:

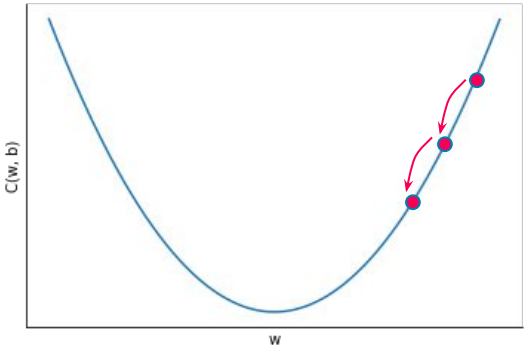


Vérité:
[0, 0, 0, 0, 0, 0, 1, 0, 0, 0]

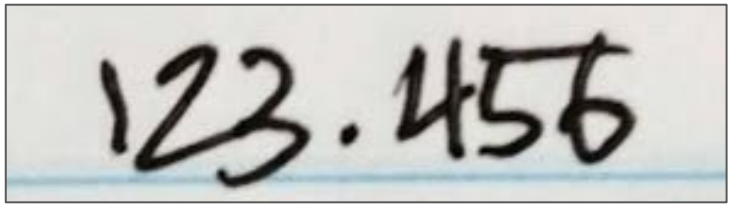


Prédictions et fonction de coût:

[0, 0, .2, 0, 0, .3, .5, 0, 0, 0]



Séquences manuscrites: formulation du problème



“123.456”

Chiffres 0-9 + séparateur décimal

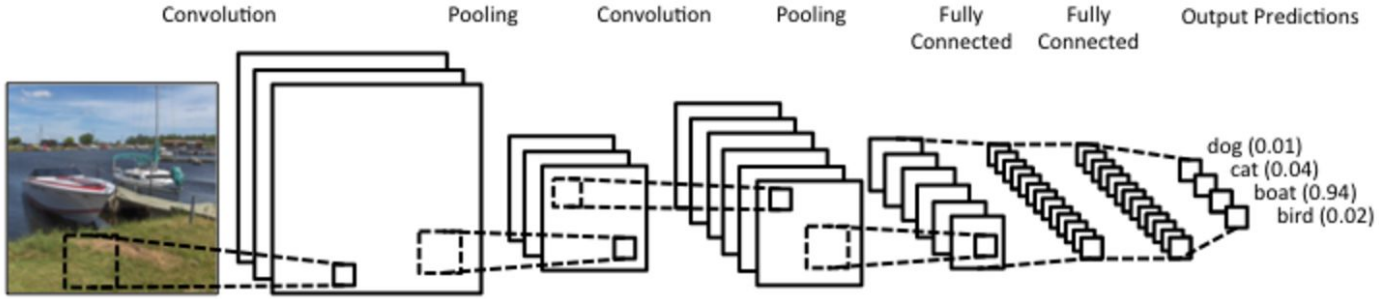
Partie entière et décimale de longueur inconnue

Largeur et position des caractères inconnues

Superposition possible entre les caractères

Approche sélectionnée: Modèle de type **CRNN** (arXiv:1507.05717)

Ingrédient 1: Réseaux de neurones convolutionnels (CNNs)



Pour les problèmes de vision numérique.

Utilise la cohérence spatiale pour réduire les connexions.

Apprentissage de *kernels* de convolution.

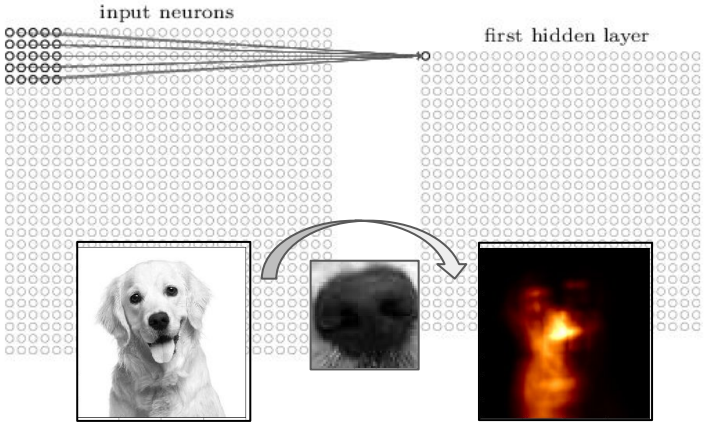
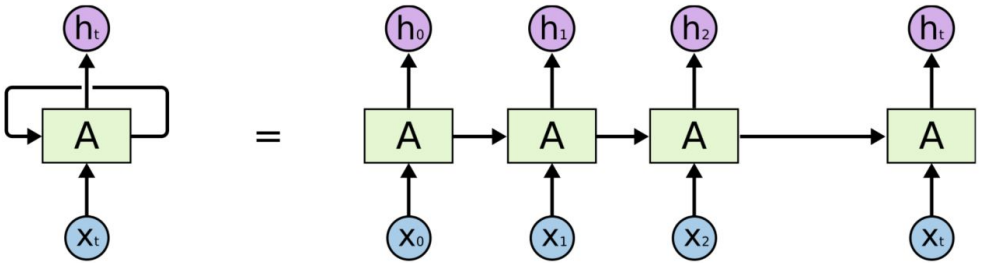


Image: https://medium.com/@Aj.Cheng/convolutional-neural-network-d9f69e473feb

Ingrédient 2: Réseaux de neurones récurrents (RNNs)



Pour les problèmes de prédiction de séquences.

Peut être vu comme une succession de copies du même réseau, chacun passant un “message” à son successeur.

État interne pour l’apprentissage de corrélations temporelles

Image: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

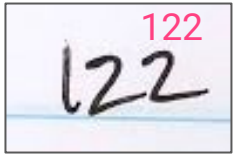
Ingrédient 3: Transcription CTC

Méthode d'encodage / décodage de séquences

Caractère immatériel de séparation "-" pour distinguer les caractères identiques adjacents.

Décodage en 2 étapes:

1. Combinaison des caractères identiques adjacents.
2. Retrait des "-" pour obtenir le mot final.



Exemple de décodage:

"---1111--222-22222"

⇩

"-1-2-2-"

⇩

"122"

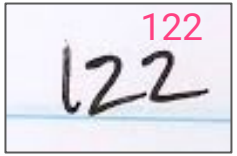
Ingrédient 3: Transcription CTC

Méthode d'encodage / décodage de séquences

Caractère immatériel de séparation "-" pour distinguer les caractères identiques adjacents.

Décodage en 2 étapes:

1. Combinaison des caractères identiques adjacents.
2. Retrait des "-" pour obtenir le mot final.



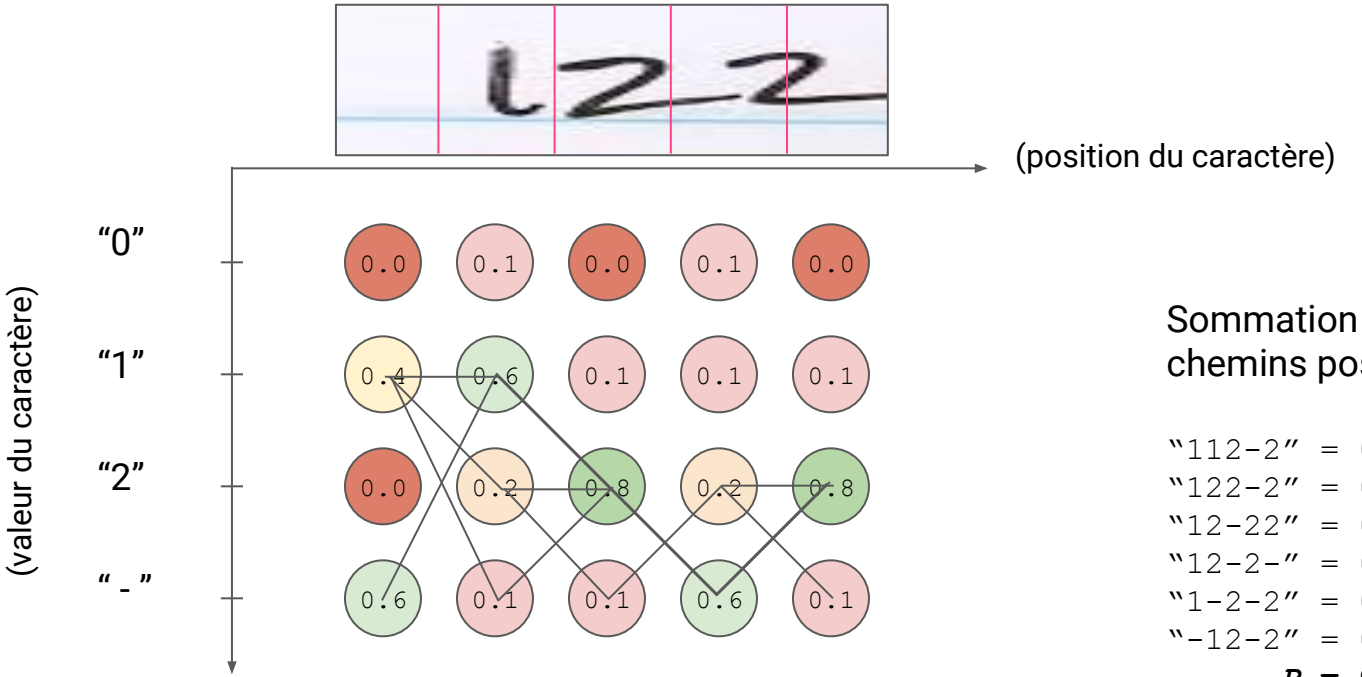
Encodages valides:

"12-2", "1122-22", "--1--2---2"

Encodages invalides:

"122", "1-22", "112222"

CTC: Calcul des probabilités et fonction de coût

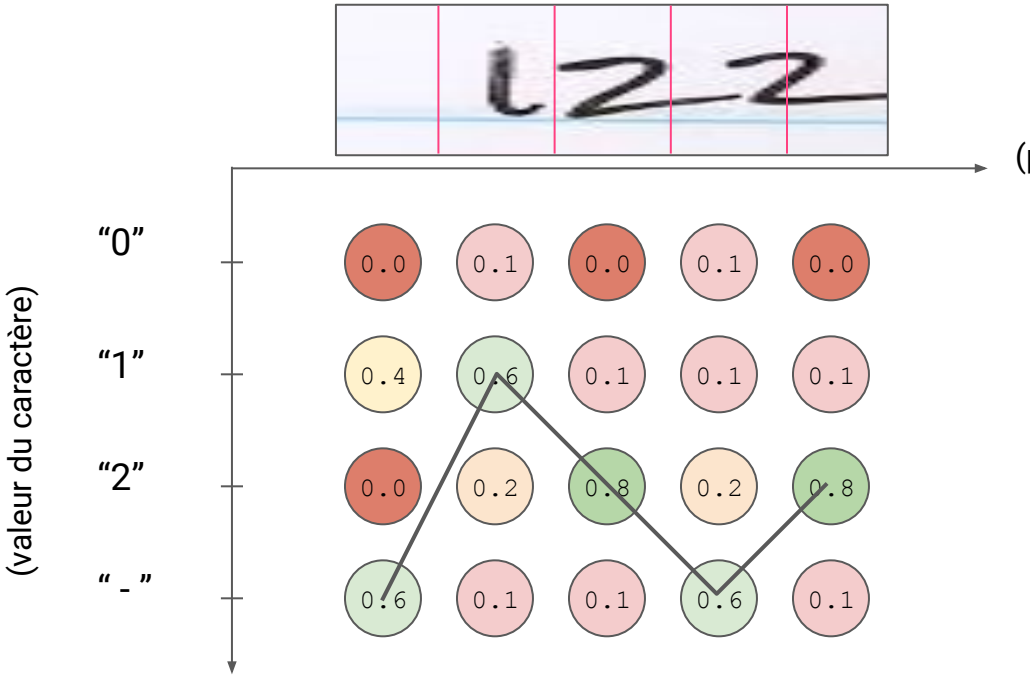


Sommation de tous les chemins possibles:

```
“112-2” = 0.0922
“122-2” = 0.0307
“12-22” = 0.0013
“12-2-” = 0.0002
“1-2-2” = 0.0154
“-12-2” = 0.1382
      P = 0.2780
```

Prédiction = distribution de probabilités

CTC: Calcul des probabilités et fonction de coût

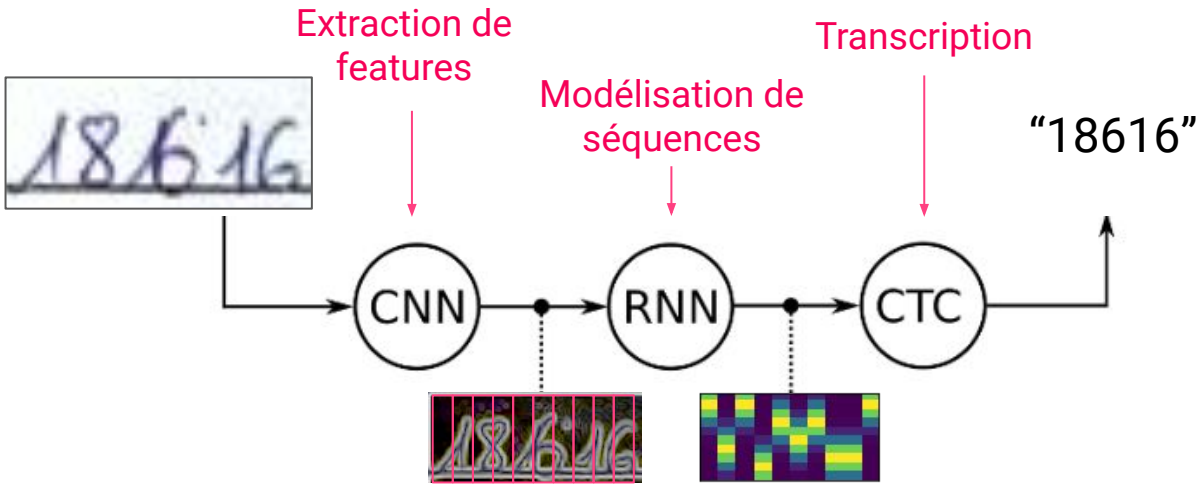


Prédiction finale

“-12-2” ⇨ “122”

Prédiction = distribution de probabilités

1 + 2 + 3: Architecture CRNN



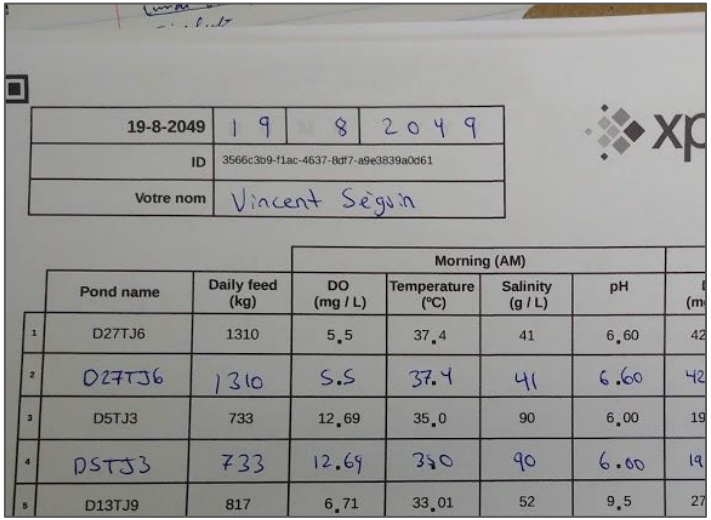
Modèle entraînable *end-to-end*

Pas nécessaire de segmenter les caractères

Facile de quantifier l’incertitude de la prédiction

B. Shi, X. Bai, C. Yao, "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition", arXiv:1507.05717.

Entraînement et données synthétiques

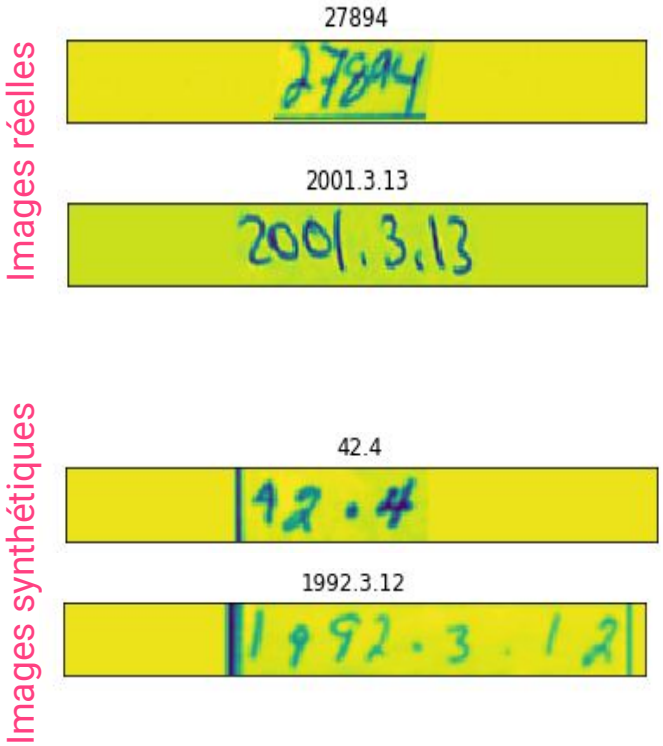


19-8-2049	19	8	2049
ID	3566c3b9-f1ac-4637-8df7-a9e3839a0d61		
Votre nom	Vincent Ségu		

	Pond name	Daily feed (kg)	Morning (AM) DO (mg / L)	Temperature (°C)	Salinity (g / L)	pH
1	D27TJ6	1310	5.5	37.4	41	6.60
2	D27TJ6	1310	5.5	37.4	41	6.60
3	D5TJ3	733	12.69	35.0	90	6.00
4	D5TJ3	733	12.69	35.0	90	6.00
5	D13TJ9	817	6.71	33.01	52	9.5

1. Formulaires “pré-annotés”

2. Génération d’images synthétiques (avec MNIST)



Y. LeCun et al., “The MNIST Dataset Of Handwritten Digits” (1999)

Merci

xpertsea.com

github.com/vmarceau

vincent@xpertsea.com