

Reconnaissance de la langue des signes en utilisant l'apprentissage profond

A dual-cam first-person vision translation system

Plan de la présentation

- Le problème
- Système de vision
- Génération des données
- Le modèle d'apprentissage profond
- Etude des résultats
- Conclusion
- Améliorations ?

Le problème

Identifier et reconnaître la
langue des signes

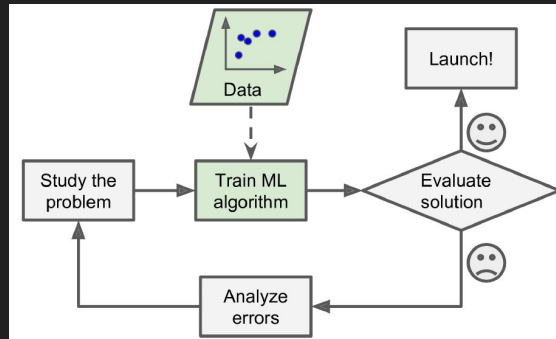
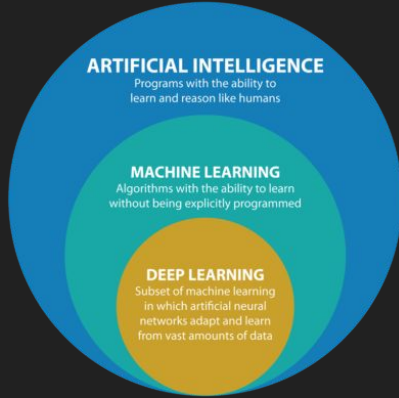
Utilisation de
l'apprentissage profond

Générer des données

Réseaux de neurones

Système
de vision

Systèmes basés
sur les contacts



Système de vision

Système de vision



Nécessité d'utiliser
plusieurs angles de vue

Pourquoi ?

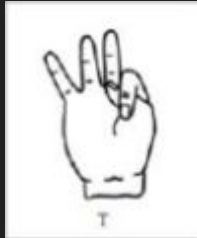


Lever l'ambiguïté
entre certains signes

Utilisation de deux caméras

Frontale

Torse

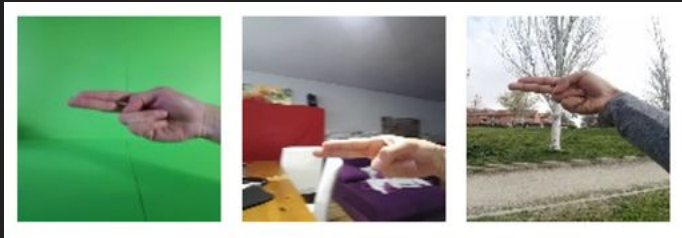


Génération des données

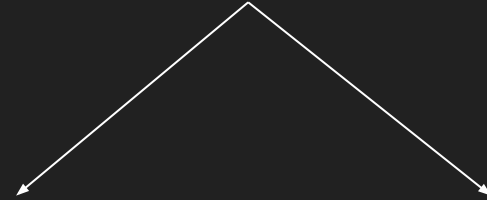
Images pour initier
le système



Demander à 4
personnes de faire
les signes

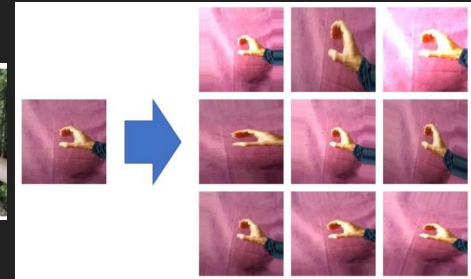


Augmenter le nombre
de données



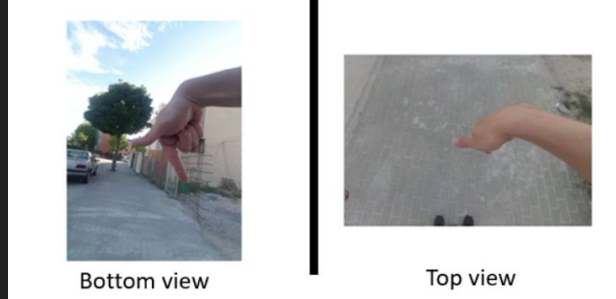
Utilisation fond vert

Distorsion



Génération des données

Vu d'un signe capturé sous les différents angles



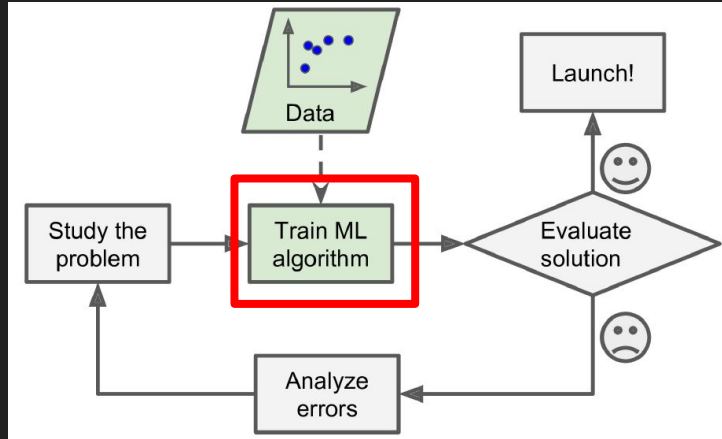
Lettre Q



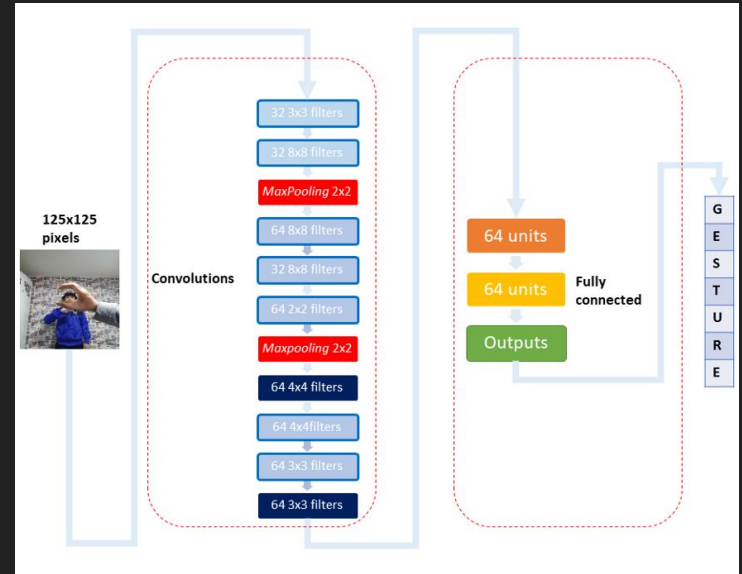
Model	Top View	Bottom View
Classes	B, C, D, E, F, G, H, I, L, O, P, Q, R, T, U, W, X	A, K, M, N, S, V, Y
Total	17	7

Problème multi-classes

Le modèle d'apprentissage profond

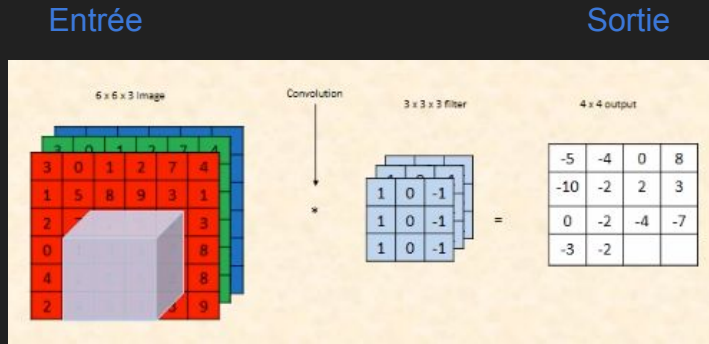


Fonctionnement de l'apprentissage profond



Modèle utilisé

Le modèle d'apprentissage profond



Convolutions

Entrée

Sortie

Max Pooling

29	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

100	184
12	45

Average Pooling

31	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

36	80
12	15

Pooling

Etude des résultats

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

$$\text{Precision} = \frac{TP}{TP + FP}$$

Taux d'identification

$$\text{Recall} = \frac{TP}{TP + FN}$$

Proportion d'images correctement reconnues
par rapport au nombre d'images proposées

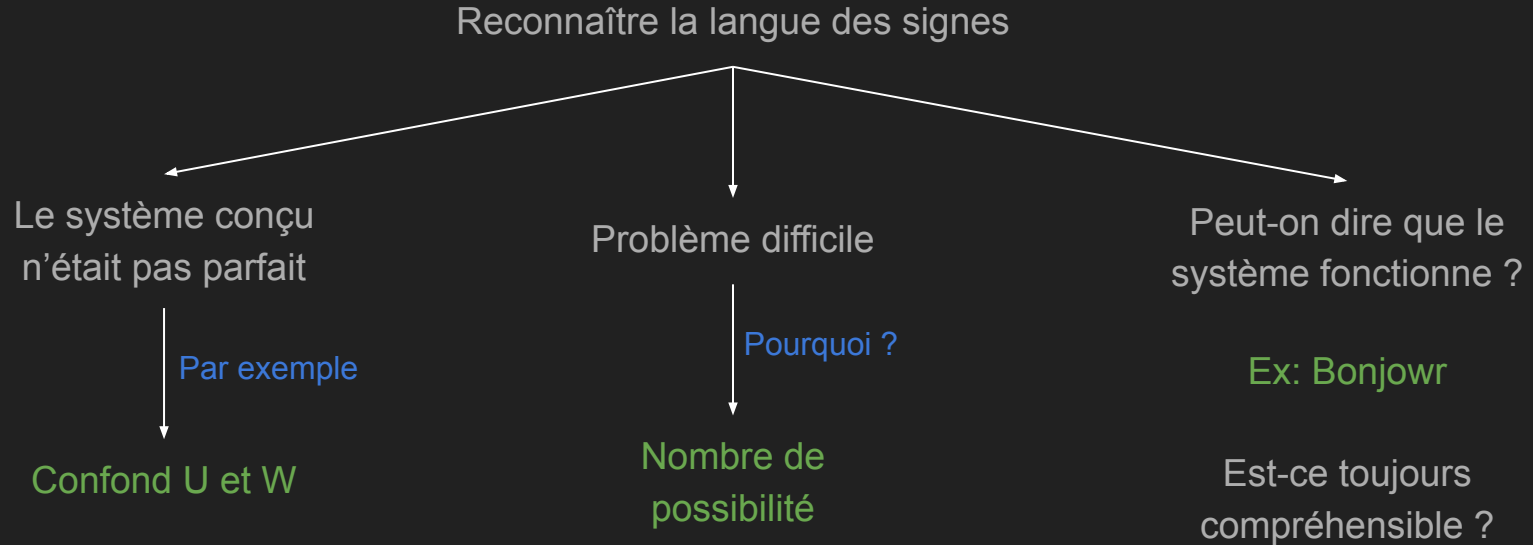
Bottom view model

	Precision	Recall
A	0,79	1,00
K	0,48	0,73
M	0,54	1,00
N	0,28	0,53
S	0,00	0,00
V	1,00	0,27
Y	0,00	0,00

Top view model

	Precision	Recall
B	0,69	0,73
C	0,39	1,00
D	0,50	0,73
E	0,52	1,00
F	1,00	0,65
G	0,65	1,00
H	1,00	0,67
I	0,67	0,65
L	0,65	1,00
O	0,67	0,93
P	0,00	0,00
Q	1,00	1,00
R	1,00	0,80
T	1,00	1,00
U	1,00	0,60
W	0,88	1,00
X	0,00	0,00

Conclusion



Améliorations ?

Comment améliorer la solution ?



José Francisco Herazo Bravo • 11:54

yeah, there are some possible improvements. I try model it as a image recognition problem and that's why I use CNNs to classify the complete image. But it is possible to model it as a **object detection problem** and use other algorithms such as Fast RCNNs, RetinaNet, YOLO, etc. I think that will probably improve the performance, the only detail is that we need to label every image again and define the area in which we find the gesture

Utiliser un autre type d'architecture d'algorithme

Si on récoltait des données qui varient davantage et utiliser le même algorithme ?



José Francisco Herazo Bravo • 10:12

oh yeah, your comment is good, I change the background in order to add complexity to my data, but yeah! What you suggest would be a better approach since you will collect more real data instead or using artificial generated ones. That should good pretty good

Ca pourrait fonctionner !

Références

La source de l'article utilisé est la suivante:

Herazo, J. (1^{er} Août 2020). Sign language recognition using deep learning : *A dual-cam first-person vision translation system*. Towards Data Science, repéré à <https://towardsdatascience.com/sign-language-recognition-using-deep-learning-6549268c60bd> (dernière consultation le 18/11/2021 - 2H38 PM EST)

Vous pourrez retrouver sur la page le lien d'une vidéo youtube de démonstration de l'algorithme ainsi que la thèse de l'auteur sur le sujet (en Espagnol) :

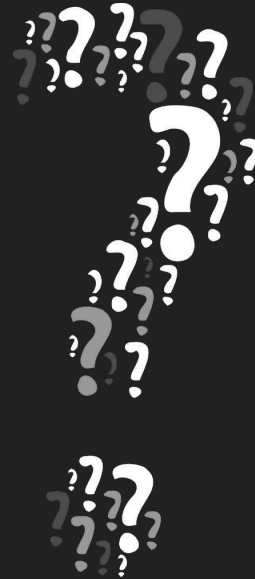
Herazo, J. (2 Août 2020). *Sign language recognition with convolutional neural networks* [Fichier vidéo]. Repéré à <https://youtu.be/pGZ7fPvnU6Q>

Herazo, J. (2 Juin 2020). Reconocimiento de señas de la lengua de señas panameña mediante aprendizaje profundo. GitHub, repéré à <https://github.com/joseherazo04/SLR-CNN/blob/master/Jos%C3%A9%20Herazo%20TFM.pdf>

Afin de comprendre le fonctionnement de l'intelligence artificielle et du deep learning, plusieurs sites ont été utilisés:

- <https://en.wikipedia.org/wiki/Convolution> (fonctionnement des convolutions)
- https://fr.wikipedia.org/wiki/R%C3%A9seau_neuronal_convolutif (réseaux de neurones)
- <https://gaussian37.github.io/dl-concept-cnn/>
- https://fr.wikipedia.org/wiki/Pr%C3%A9cision_et_rappel
- <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall?hl=fr>

Merci de votre attention



Avez-vous des questions ?

