SCIENCES SORBONNE UNIVERSITÉ

Commande Vocale

3 juin 2024

Saliou BARRY & Carine Moubarak

Sous la supervision de Mr. Olivier SCHWANDER

Introduction

“Hey Siri”

“Hey Cortana”

“Alexa”

“OK Google”

“Hi Bixby”

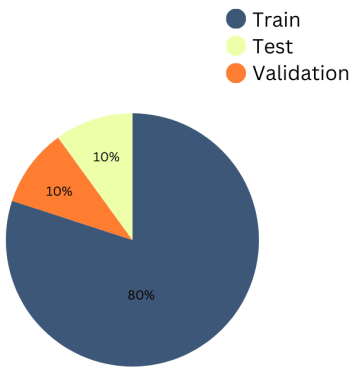
DATASET et Extraction

Dataset

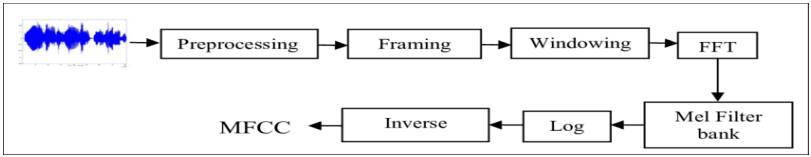
SPEECHCOMMAND v0.1 (Pytorch)

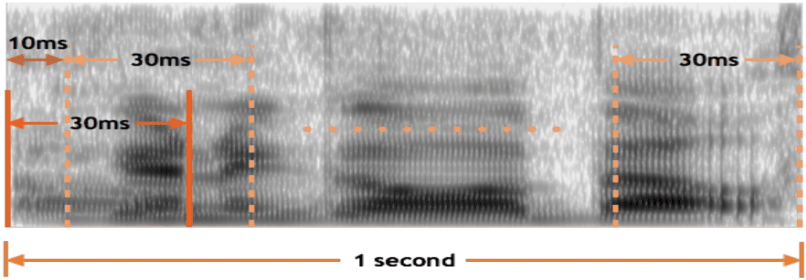
1. Taille : 64727 audios
2. Durée : 1s
3. Label : 30 (ex : ”Yes”,”No”,”Go”, ...)

Partition des données



Extraction MFCC





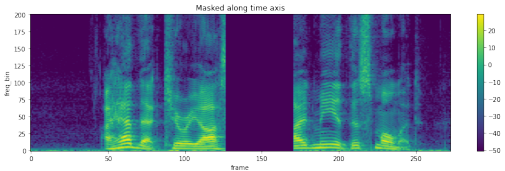
Augmentation

- Time Shift : [-100,100]
- Background Noise : [0.05,0.3]

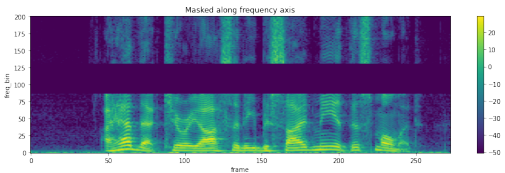
Stratégie d'augmentation

- Random : 40%
- Duplication

Time Masking : [0,25]



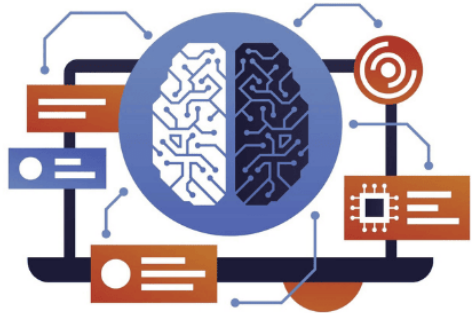
Frequency Masking : [0,15]



Modèles

Modèles

1. Régression Logistique
2. Simple Neural Network
3. Convolutional Neural Network

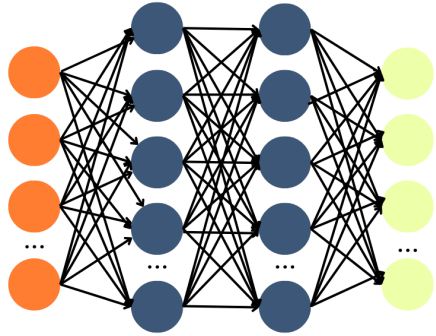


Régression Logistique

RL	Acc Validation	Acc Test	Type
No Augment	31.9%	30.3%	-
Time Shift	32%	30.2%	random
Time Masking	31.7%	30%	random
Frequency Mask	30%	29.6%	random
Background Noise	33%	30.3%	random
Combined Aug	31.9%	30%	random

Table – Résultats d'expérimentation

Simple Neural Network

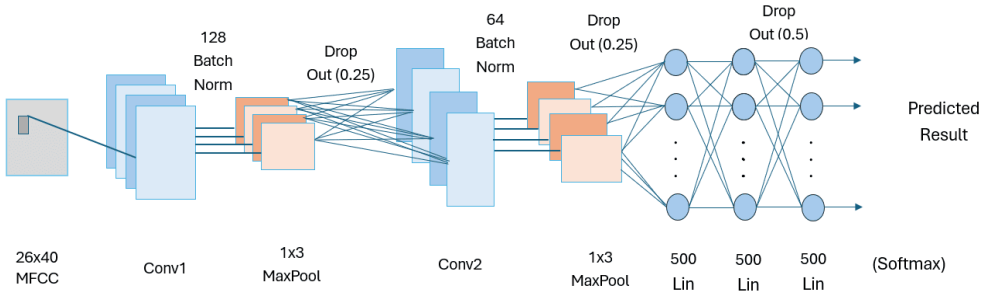


Simple Neural Network

SNN	Acc Validation	Acc Test	Type
No Augment	73.7%	73%	-
Time Shift	76.9%	76.2%	random
Time Masking	75.6%	75%	random
Frequency Mask	74.6%	74.1%	random
Background Noise	75.9%	74.5%	random
Combined Aug	76.6%	75.4%	random

Table – Résultats d'expérimentation

Convolutional Neural Network



Convolutional Neural Network

CNN	Acc Validation	Acc Test	Type
No Augment	90.1%	90%	-
Time Shift	90.5%	90.8%	random
Frequency Mask	87.8	87	random
Background Noise	88.3%	87.2%	random
Time Masking	89.2	88.7	random
Combined Aug	90.4%	90%	random

Table – Résultats d'expérimentation

Conclusion

Récapitulatif des performances de chaque modèle

	CNN	SNN	LR
Accuracy	90,8%	74.5%	30,1%
Precision	91%	75.4	34,7%
Recall	90,8%	75.1	30,5%
F1-Score	91%	74.9	30,1%

	Acc	class	Dataset
AST	98.11%	35	v0.2
KWS	95%	12	v0.1

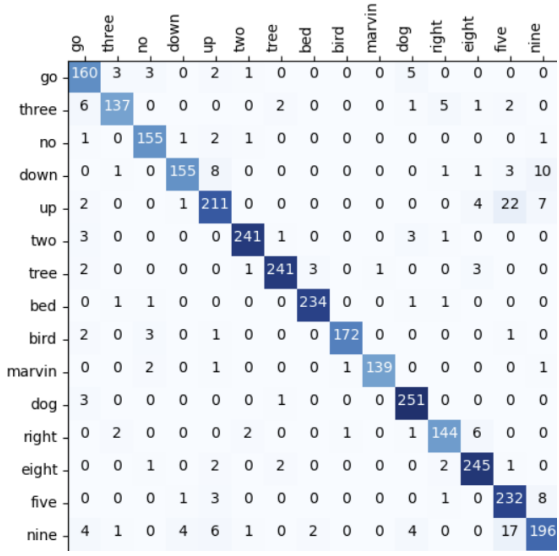
SCIENCES
SORBONNE
UNIVERSITÉ

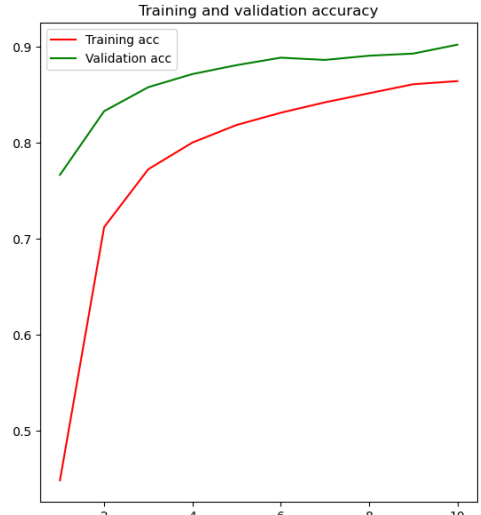
Thank you for your attention

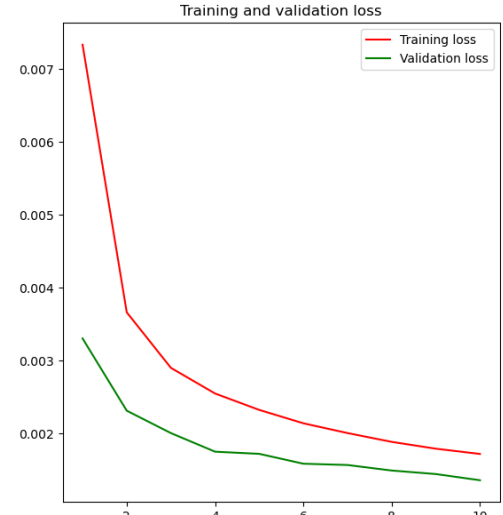
Annexe

Table – Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
go	0.78	0.78	0.78	251
no	0.80	0.92	0.86	252
bed	0.84	0.91	0.87	176
up	0.81	0.93	0.86	272
three	0.85	0.86	0.86	267
tree	0.91	0.84	0.87	193
...	...	...	...	...
Précision globale			0.91	6835
Moyenne pondérée	0.91	0.91	0.91	6835
weighted avg	0.91	0.91	0.91	6835


go three no down up two tree bed bird marvin dog right eight five nine
go 160 3 3 0 2 1 0 0 0 0 5 0 0 0 0
three 6 137 0 0 0 0 2 0 0 0 1 5 1 2 0
no 1 0 155 1 2 1 0 0 0 0 0 0 0 0 1
down 0 1 0 155 8 0 0 0 0 0 0 1 1 3 10
up 2 0 0 1 211 0 0 0 0 0 0 0 4 22 7
two 3 0 0 0 0 241 1 0 0 0 3 1 0 0 0
tree 2 0 0 0 0 1 241 3 0 1 0 0 3 0 0
bed 0 1 1 0 0 0 0 234 0 0 1 1 0 0 0
bird 2 0 3 0 1 0 0 0 172 0 0 0 0 1 0
marvin 0 0 2 0 1 0 0 0 1 139 0 0 0 0 1
dog 3 0 0 0 0 0 1 0 0 0 251 0 0 0 0
right 0 2 0 0 0 2 0 0 1 0 1 144 6 0 0
eight 0 0 1 0 2 0 2 0 0 0 0 2 245 1 0
five 0 0 0 1 3 0 0 0 0 0 0 1 0 232 8
nine 4 1 0 4 6 1 0 2 0 0 4 0 0 17 196


Training and validation accuracy
Training acc
Validation acc
0.9
0.8
0.7
0.6
0.5
2
4
6
8
10


Training and validation loss
Training loss
Validation loss
0.007
0.006
0.005
0.004
0.003
0.002
2
4
6
8
10