Apprentissage automatique pour la reconnaissance d'émotions et de styles à partir de mouvements humain.

Tuteur : M. Maxime Devanne

Par Melchior Obame Obame & Saad Bendaoud
3A IR

Plan

I. Rappel sur la première présentation

   Objectifs, K-NN, Bases de données

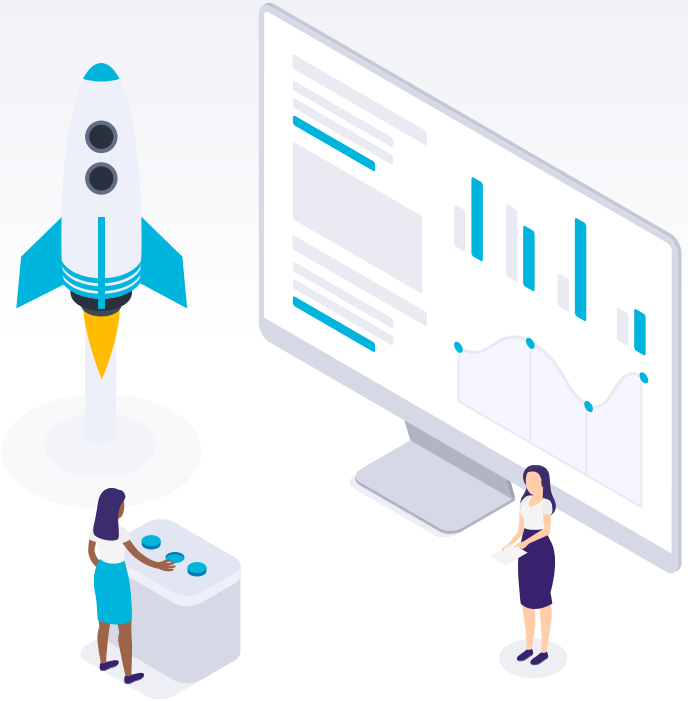
II. Evolutions du projet

   1. LSTM
   2. CNN

III. Conclusion

   1. Problèmes rencontrés
   2. Optimisations
   3. Conclusions finales

I Rappels



Objectifs

- Construire un modèle capable de prédire le sujet qui réalise une action.
  - Ici on identifie le sujet et non l'action.

Méthodes d'apprentissage

- Algorithmes d'apprentissage utilisés :
  - Apprentissage supervisé **K-NN** (k-nearest neighbors).
  - **CNN** : Convolutional Neural Network.
  - **RNN** (**LSTM**) : Recurrent Neural Network .

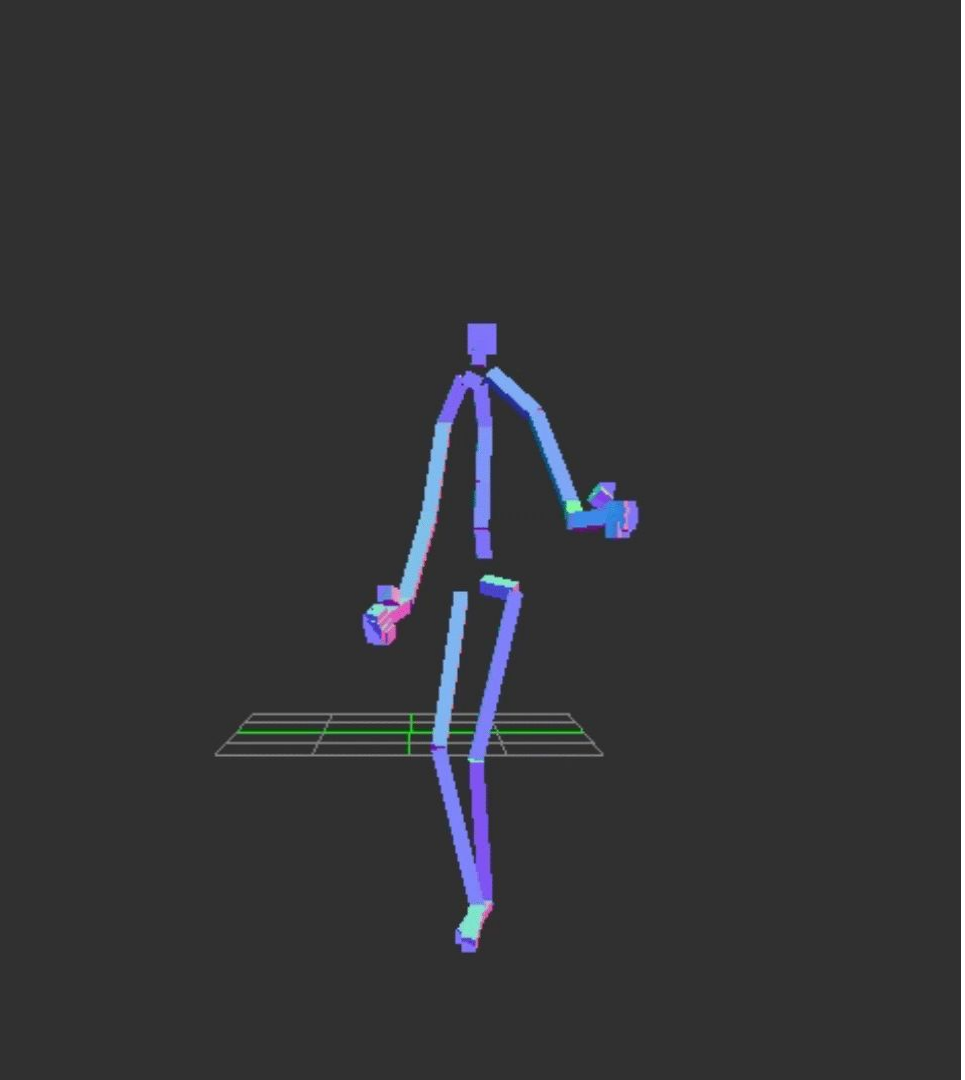
KNN

Présentation des bases de données :

	Séquences	Sujets	Classes	Joints	Train |Test %
MsrAction3D	567	10	20	20	66.3 | 33.7
Emotional Body Motion	1447	8	11	23	75.3 | 24.7
Dance Motion Capture	134	9	12	54	69.2 | 30.8

Emotional Body Motion DB

Tristesse



Dance motion capture DB

Flamenco

Mise en place du KNN

Preprocess

Enregistrement des données sous forme de fichier CSV

Importer la DataSet

Transformation de la dataSet en numpy array d'actions.

Mise en place K-NN

Fixation du nombre de voisins optimal ainsi que des hyperparamètres.

Prédiction

Prédiction des données de test.

KNN

Rappels:

Consiste à prendre en compte les k échantillons d'apprentissage dont l'entrée est la plus proche d'une action a, selon une distance à définir.

Distance utilisée: **DTW**

Soient F_k et F_p deux frames constituées de joints tel que $F_k = \{J_k^1, J_k^2, \ldots, J_k^N\}$ et $F_p = \{J_p^1, J_p^2, \ldots, J_p^N\}$ où N représente le nombre de joints .

Notons DJ_{ik} la distance entre deux joints J_i et J_k et DF_{kp} *la distance entre les deux frames k et p, On a :*

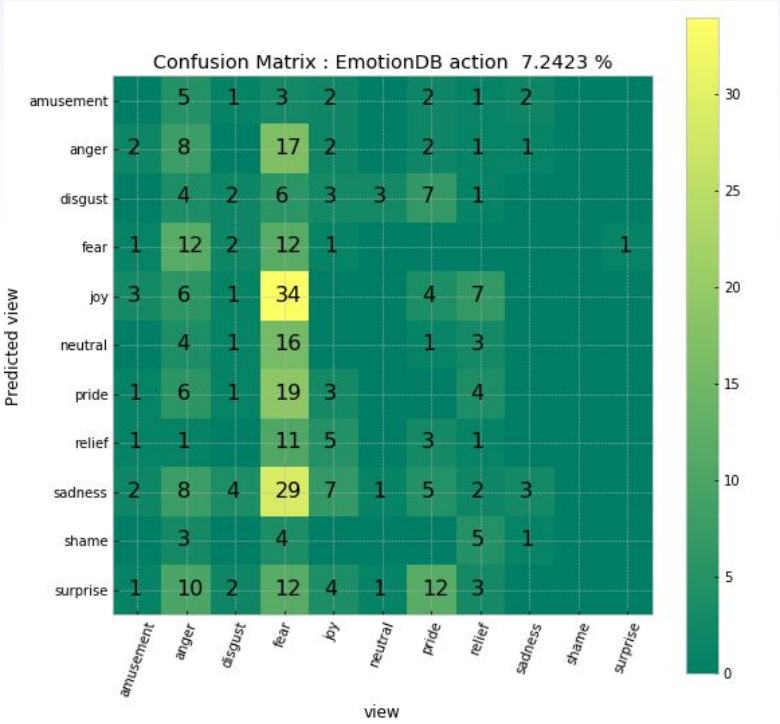
$$DJ_{ik} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2} \quad (1)$$

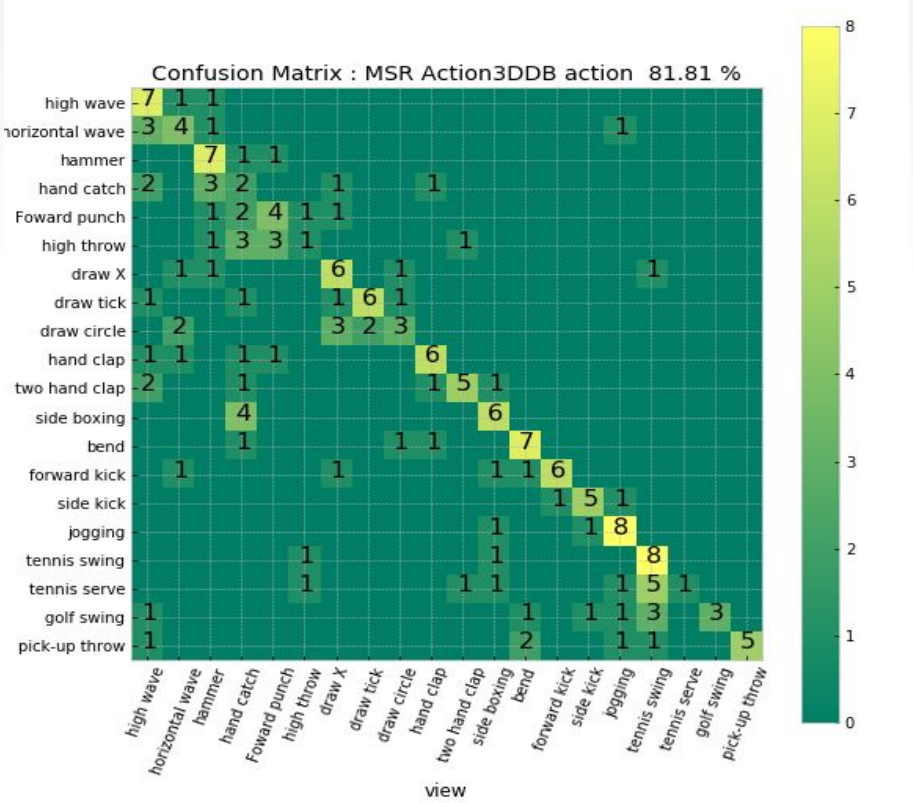
$$DF_{kp} = \sum_{i=1}^{N} DJ_{kp}^i \quad (2)$$

Résultats du KNN

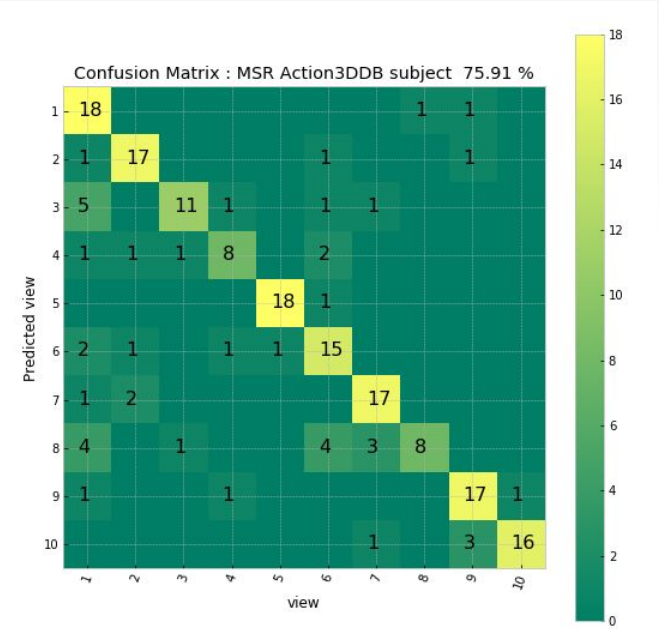
	KNN Accuracy Action	KNN accuracy Sujet
MSR Action 3D DB	**81.81%**	**75.91%**
Dance Motion DB	**8.33%**	**44.73%**
Emotional Body Motion DB	**7.24%**	**10.02%**

Résultats du KNN

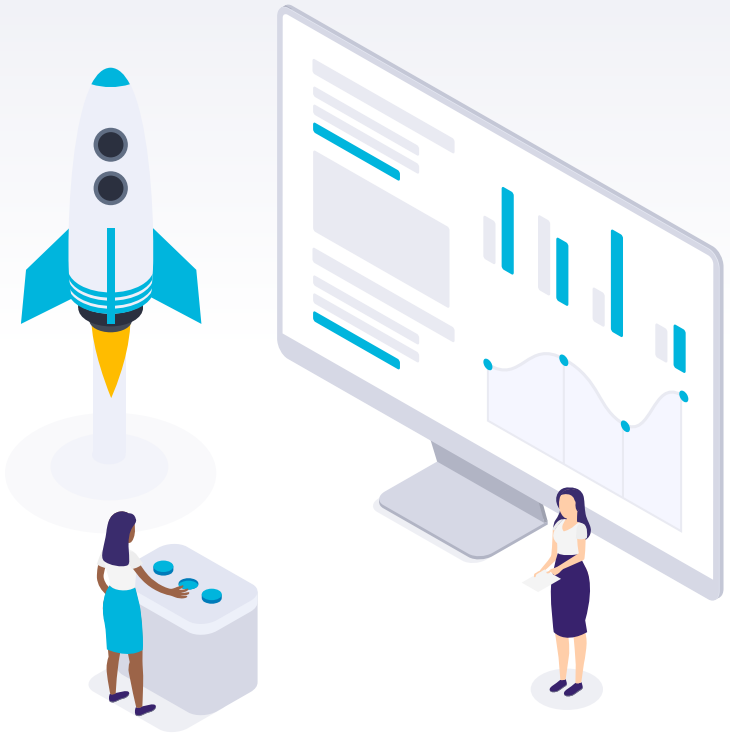




Résultats du KNN



Evolutions



Méthodes de deep learning

Pourquoi utiliser les réseaux de neurones ?

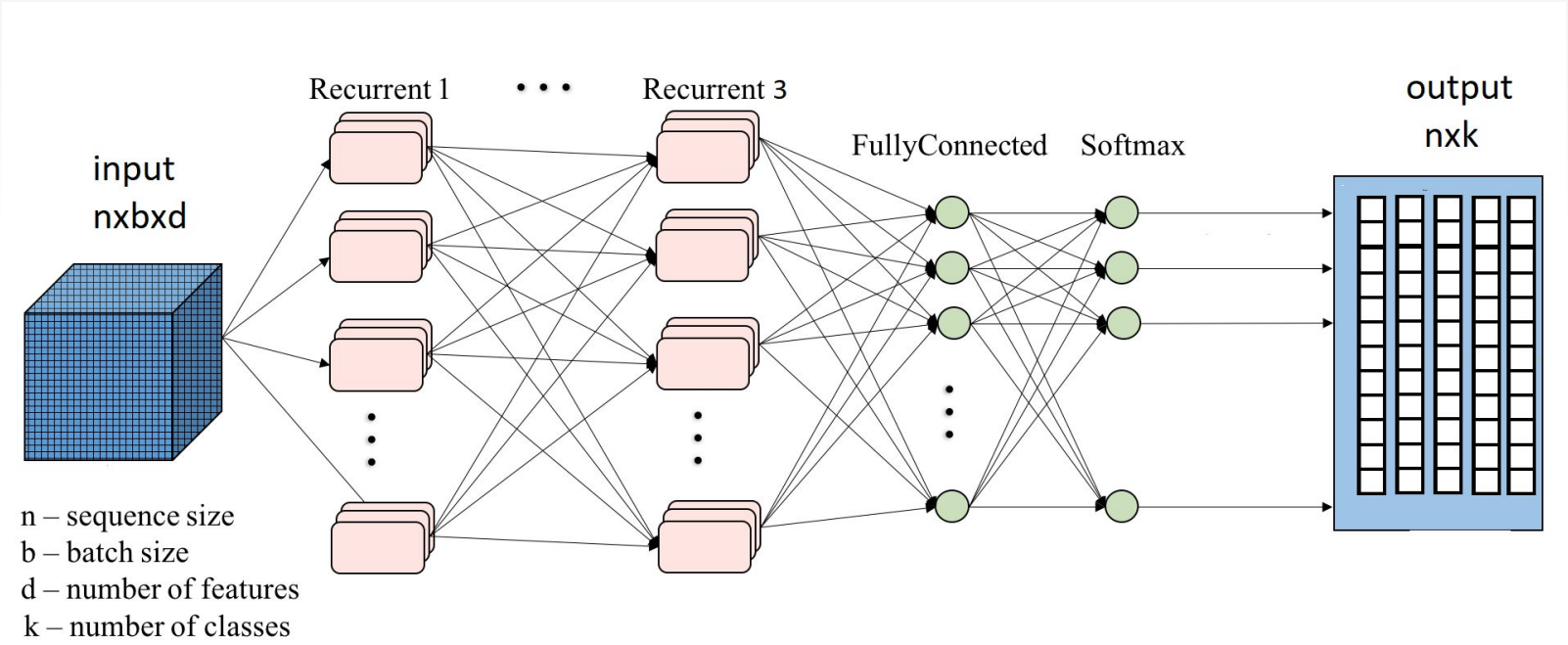
- Plus besoins de comparaison sur les données d'entraînement, une fois le réseau entraîné.
- En théorie plus précis que K-NN mais plus lent lors de l'apprentissage

LSTM

Avantages des LSTM (Long Short Term Memory) :

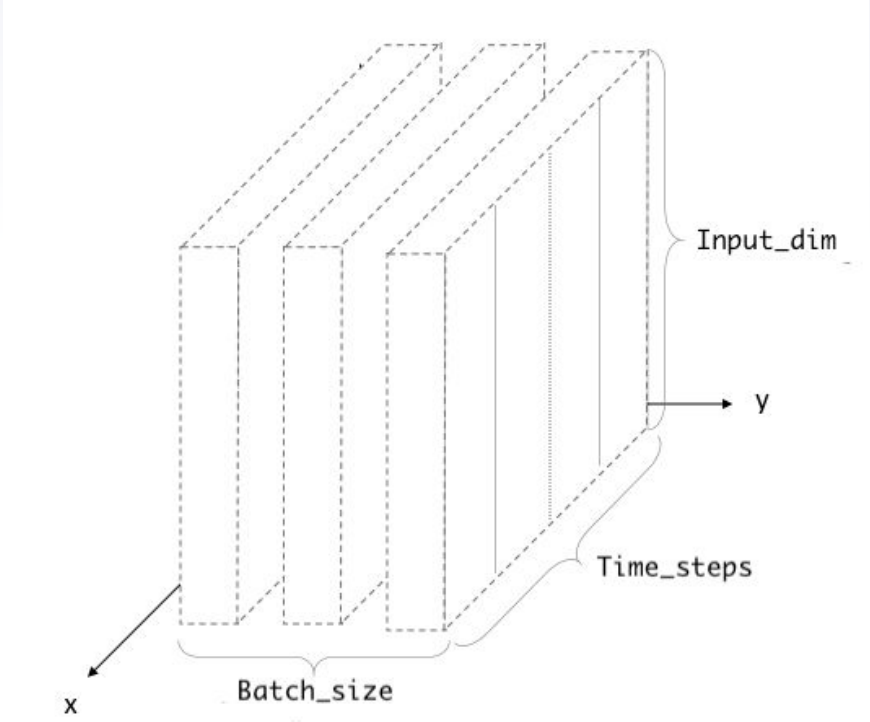
- Les LSTM sont optimisés pour travailler sur des données avec des relations temporelles.
- Plus besoin de DTW.

LSTM


Recurrent 1
Recurrent 3
output
nxk
input
nxbxd
FullyConnected
Softmax
n – sequence size
b – batch size
d – number of features
k – number of classes

LSTM

L'input shape du réseau LSTM



LSTM

Choix des hyper paramètres:

- Nombre de neurones: correspond aux nombre de features (=nombre d'articulation x son nombre de coordonées)
- Loss function: categorical cross entropy

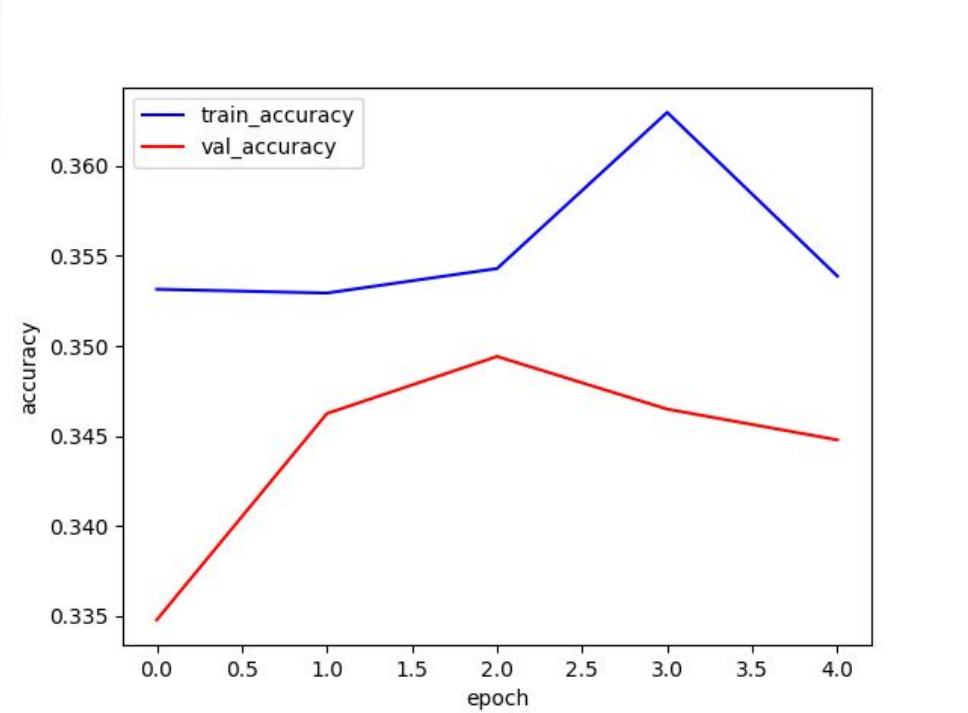
LSTM

Resultats LSTM :

	LSTM accuracy Sujet
MSR Action 3D DB	**37.01%**

LSTM

la variation d'accuracy en fonction du nombre d'époque pour MSRAction3D



CNN

Avantages des CNN (Convolutional Neural Network):

- Les CNN sont optimisés pour les traitements d'images

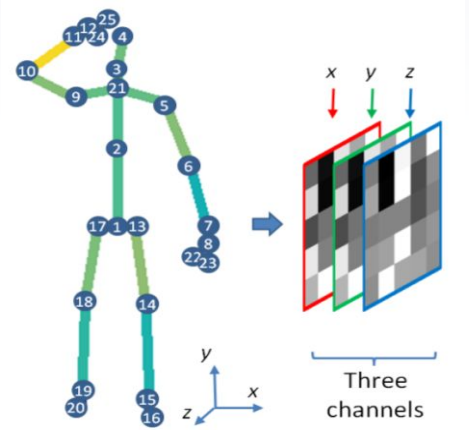
CNN

Méthode d'obtention des images à partir des squelettes.

$$x_i' = 255 \cdot \frac{x_i - \min\{C\}}{\max\{C\} - \min\{C\}}$$

$$y_i' = 255 \cdot \frac{y_i - \min\{C\}}{\max\{C\} - \min\{C\}}$$

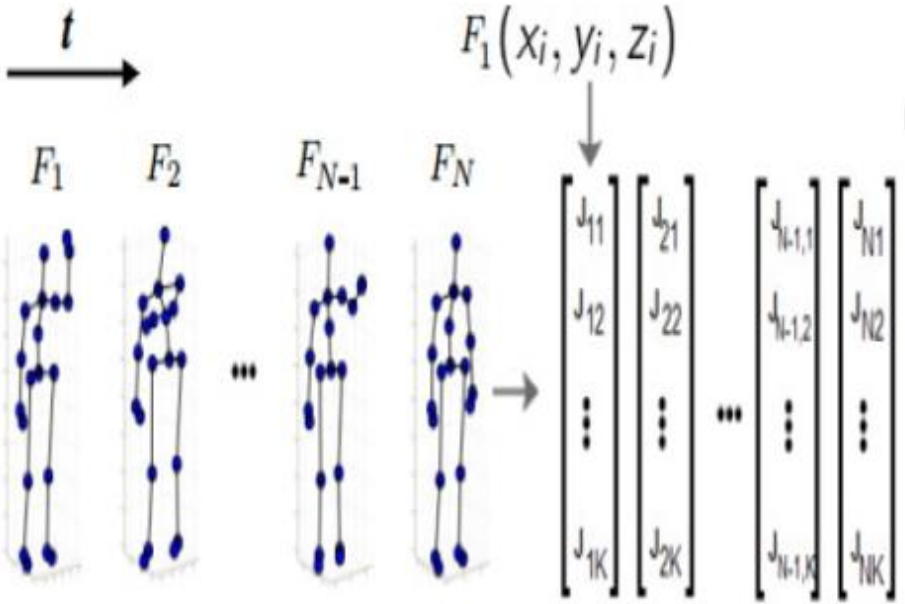
$$z_i' = 255 \cdot \frac{z_i - \min\{C\}}{\max\{C\} - \min\{C\}}$$

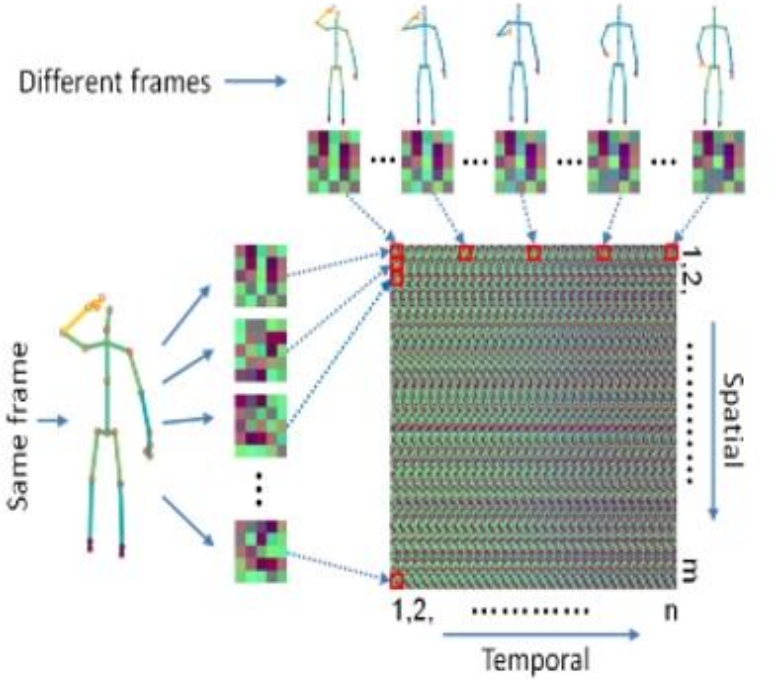


les coordonnées 3D (x_i, y_i, z_i) de chaque squelette $\{F_k\}$

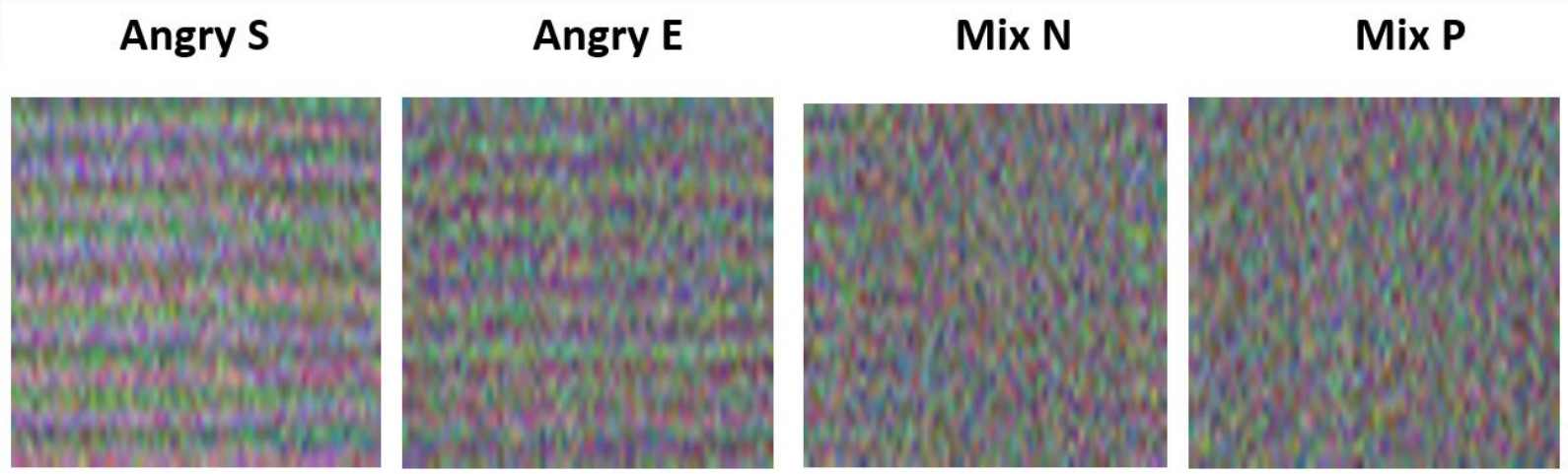
CNN

Méthode d'obtention d'images à partir des séries temporelles



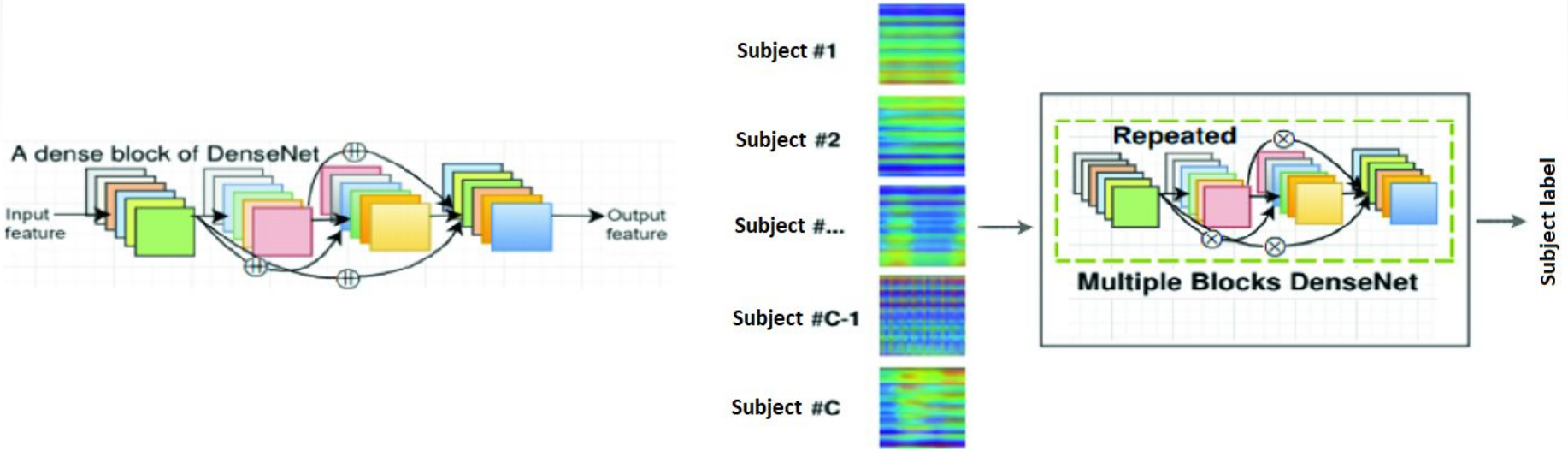


CNN



CNN

Architecture choisi pour le CNN



CNN

Notre Modèle d'architecture CNN:

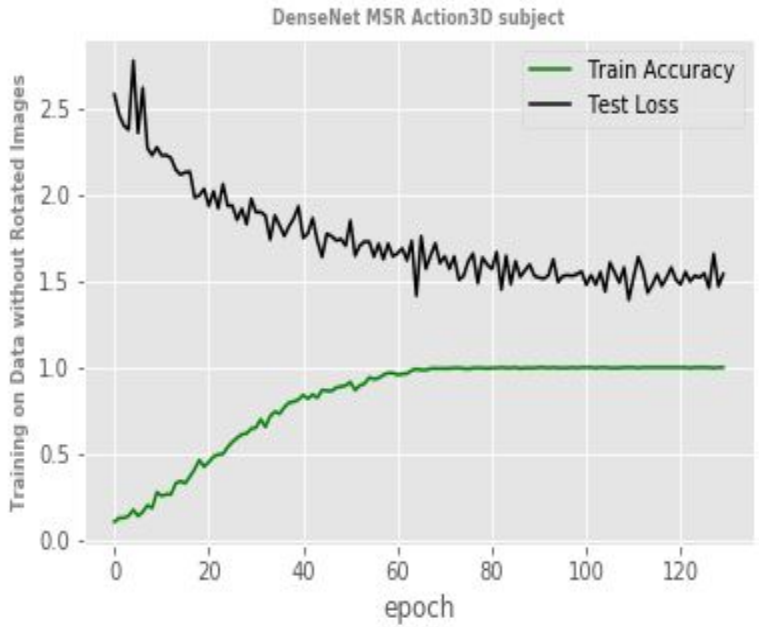
Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 30, 30, 32)	896
max_pooling2d_3 (MaxPooling2	(None, 15, 15, 32)	0
conv2d_4 (Conv2D)	(None, 13, 13, 64)	18496
max_pooling2d_4 (MaxPooling2	(None, 6, 6, 64)	0
flatten_2 (Flatten)	(None, 2304)	0
dense_3 (Dense)	(None, 256)	590080
dense_4 (Dense)	(None, 20)	5140

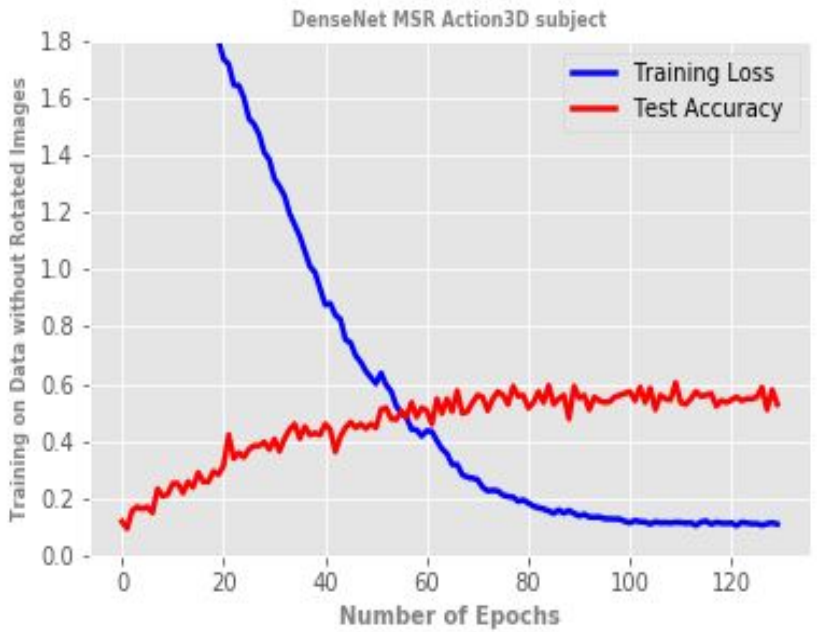
CNN

Resultats CNN :

	CNN Accuracy Subject
MSR Action 3D DB	**60%**
Dance Motion DB	**10.23%**
Emotional Body Motion DB	**50.7%**

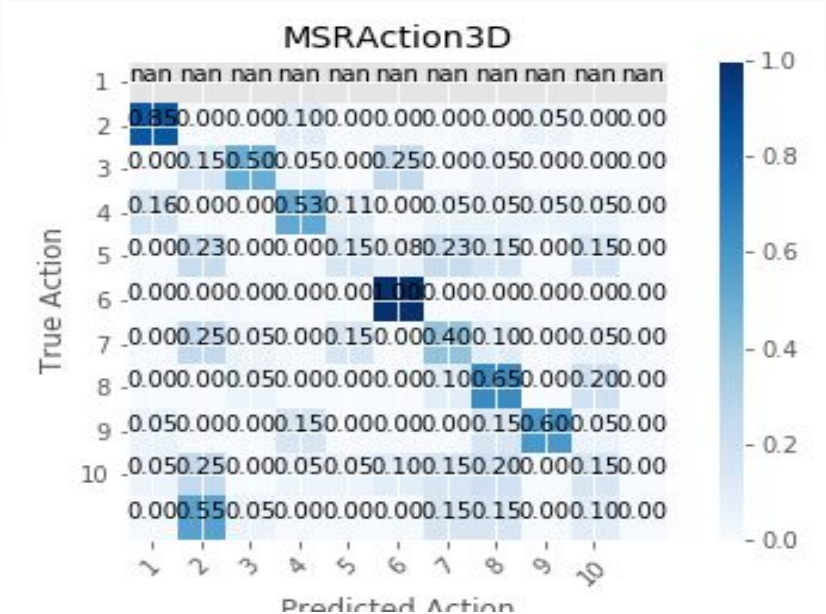
CNN


DenseNet MSR Action3D subject
Train Accuracy
Test Loss
Training on Data without Rotated Images
epoch
0.0
0.5
1.0
1.5
2.0
2.5
0
20
40
60
80
100
120


DenseNet MSR Action3D subject
Training Loss
Test Accuracy
Training on Data without Rotated Images
Number of Epochs
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
0
20
40
60
80
100
120

CNN

Résultats de la prédiction des 10 sujets de MSR Action 3D :



III

Conclusion

Problèmes rencontrés

- Temps d'entraînement de ces 3 algorithmes sur chacunes des bases de données.
- Choix de la bonne architecture
- Choix des bon hyperparamètres

Optimisations

- Tester d'autres architectures (CNN+LSTM,...)
- Optimisations sur les hyperparamètres.

Conclusions

.

- Les algorithmes de Deep Learning ne sont pas toujours les plus performants.
- Nos algorithmes restent optimisables

Sources

- KNN et la DTW : https://github.com/markdregan/K-Nearest-Neighbors-with-Dynamic-Time-Warping
- LSTM et CNN : https://github.com/microsoft/View-Adaptive-Neural-Networks-for-Skeleton-based-Human-Action-Recognition
- Article MSR Action 3D : https://documents.uow.edu.au/~wanqing/#Datasets
- HOG et autres méthodes : https://www.researchgate.net/figure/MSR-Action-3D-We-compare-our-method-with-HO3DJ-2-EigenJoints-4-STOP-7-HOG_tbl1_278767979

Sources

- KNN pour les émotions : https://www.researchgate.net/figure/P-BME-dataset-Emotion-recognition-accuracy-obtained-using-a-nearest-neighbor-approach_tbl2_317870981
- ELM Dance DB : https://www.researchgate.net/figure/Each-classification-rate-obtained-by-4-fold-cross-validation_fig6_317293103
- Emotional body database : https://www.researchgate.net/publication/269174908_The_MPI_Emotional_Body_Expressions_Database_for_Narrative_Scenarios

Merci pour votre attention

Des questions ?