

Data

Cross user sessions → 2, 3, 5, 6, 8, 9, 10, 11, 12, 13 : 11 gestes x 25 instances x 10 sessions = 2750 exemples

Single user sessions → 0, 1, 4, 7, 13, 14 : 11 gestes x 50 instance x 5 sessions = 2750 exemples (reconnaissance personnalisée)

→ Classes équilibrées : 250 exemples par geste

On utilise ici les données de sessions cross user : DATA(num_geste, num_sessions, num_instance, num_channel, num_frame, 32, 32)

1 geste → DATA(num_geste, num_sessions, num_instance, :, :, :,:)

Chaque geste est constitué de 40 frames (padding pour pouvoir apprendre en batch), ce qui constitue l'échelle temporel lente. Chaque frame est un spectrogramme de 32x32 → on retrouve dans chaque frame l'échelle de temps rapide.

Chaque geste est donc un volume de 40x32x32 à 4 channels (channels hardware radar soli)

Models

Layers	CNN_3D (47 353 195 params)	RNN_CNN_3D (29 159 131 params)	SPD_net (741 387 params)	SPD_net_BN	MDM pyriemann
1	Conv1 3x3x3x32	Conv1 3x3x3x32	Cov_pool	Cov_pool	
2	Conv2 3x3x3x64	Conv2 3x3x3x64 + Maxpool3D 2x2x2	Bimap1 512	Bimap1 512 + BN	
3	Conv3 3x3x3x128 + Maxpool3D 2x2x2	Fc1 128	Bimap2 256	Bimap2 256 + BN	
4	Fc1 128	Fc2 128	Bimap3 128	Bimap3 128 + BN	
5	Fc2 11 + softmax	LSTM (hidden size : 100)	Bimap4 64	Bimap4 64 + BN	
6		Fc3 11 + softmax	LogEig1	LogEig1	
			Fc1 11 + softmax	Fc1 11 + softmax	

Résultats (5 folds cross val)

[illegible]