Extracting activity masks for each class
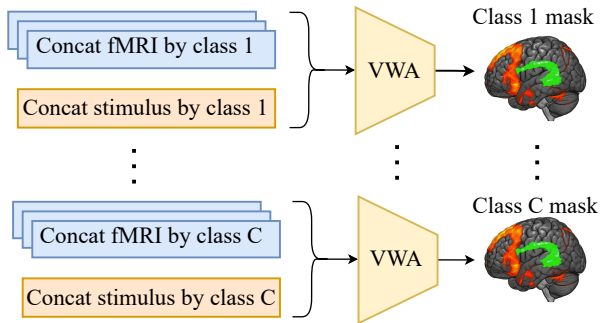
Concat fMRI by class 1
Concat stimulus by class 1
VWA → Class 1 mask

Concat fMRI by class C
Concat stimulus by class C
VWA → Class C mask

Classification

fMRI tensors
$\tau \times X \times Y \times Z$
AvgPool

Class 1 mask → $h_1 \times \tau$ → TSM
Class 2 mask → $h_2 \times \tau$ → TSM
Class C mask → $h_C \times \tau$ → TSM

$d \times 1$
Concat embeddings → Simple Classifier

Spatio-temporal feature extraction