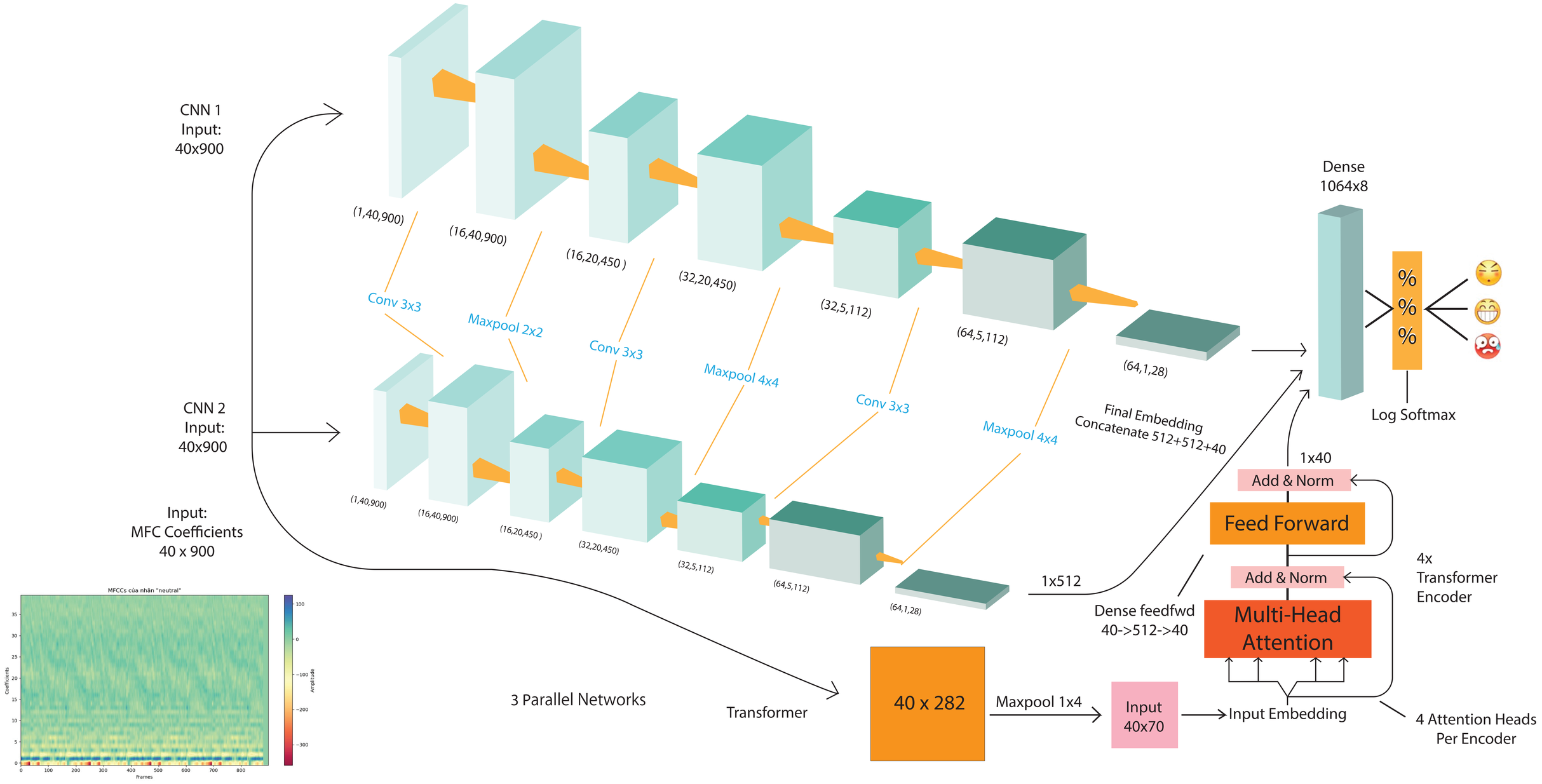
CNN 1
Input:
40x900
CNN 2
Input:
40x900
Input:
MFC Coefficients
40 x 900
(1,40,900)
(16,40,900)
(16,20,450)
(32,20,450)
(32,5,112)
(64,5,112)
(64,1,28)
Conv 3x3
Maxpool 2x2
Conv 3x3
Maxpool 4x4
Conv 3x3
Maxpool 4x4
Final Embedding
Concatenate 512+512+40
Dense
1064x8
%
%
%
Log Softmax
1x40
Add & Norm
Feed Forward
Add & Norm
Multi-Head
Attention
4x
Transformer
Encoder
1x512
Dense feedfwd
40->512->40
4 Attention Heads
Per Encoder
Input Embedding
Input
40x70
Maxpool 1x4
40 x 282
Transformer
3 Parallel Networks
MFCCs của nhãn "neutral"
Coefficients
Frames
Amplitude