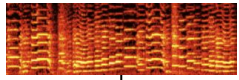

Log-mel spectrogram patches of 2s
100 frames @ 40ms (50% overlap), 96 bands

BN + **ReLu** + **Conv2D**(24, 5x5)
BN + **ReLu** + **MaxPool2D**(4x2)

BN + **ReLu** + **Conv2D**(48, 5x5)
BN + **ReLu** + **MaxPool2D**(4x2)

BN + **ReLu** + **Conv2D**(48, 5x5)
BN + **ReLu**

Flatten + **Dropout**(0.5)

Dense(64) + **ReLu** + **Dropout**(0.5)

SoftMax(20)

Trained with:
- initializer: he_normal
- SAME padding
- Adam optimizer
- batch size 64
- learning rate 10^{-3}
- halving lr when val_acc plateaus

one of the proposed loss functions

Total:
531,624 weights

Aggregate predictions over all clip patches

Clip-level prediction