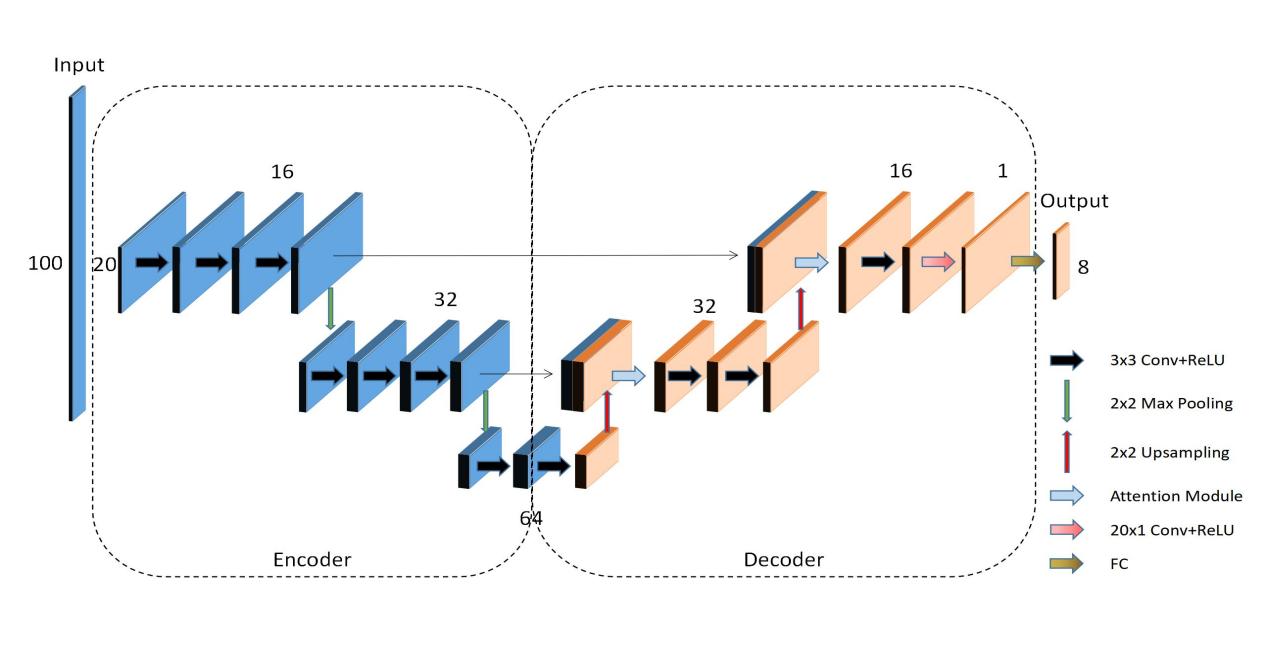
同时我们使用深度学习的方法进行了实验。经调研，在当下神经网络的使用中，相较于循环神经网络（RNN）和最新的Transformer，卷积神经网络（CNN）的用途较为广泛且成熟度较高，因此我们选择使用CNN进行对比实验。又由于在传统的CNN任务中，不同的任务适用于不同的场景，比如：图像分类比较适用于多种类（大于百类），检测任务则需要对目标进行框标记等。其中，语义分割任务则是需要将图像中所有的像素点进行简单分类（少于百类）。而鉴于此任务的性质：多组数据，较少类别的分类任务。这十分契合语义分割任务的目标，因此我们选择使用语义分割任务中较为著名的UNet[1]网络结构作为baseline进行实验，其优点是融合了多尺度信息，保证了提取特征和原始信息不会丢失，并且当信号持续时间增长时可以通过简单地加深网络深度的方法处理，而不会过多加重训练负担。又由于近年来Transformer中注意力机制的兴起，我们使用了[2]中提出的轻量化CNN全局注意力模块对网络进行了修改，整体网络结构如下图所示（以三层尺度为例）：



其中输入为1\*100的信号向量，由于雷达信号为一秒20帧，因此，将信号reshape为20x5的矩阵，以便更好地提取时间轴上的特征，并送入Encoder。

送入的信号矩阵会进行如下操作：

（1）进行3x3的卷积操作后通道变为16。

（2）进行两次3x3的卷积操作后得到20x5尺度上的特征。

（3）对其进行最大池化得到。

即

上述步骤适用于Encoder中所有中间层和首层，以此类推，就得到了10x3尺度上的32通道特征。而最后一层的处理应如图所示，仅进行一次卷积操作即可得到。

不同尺度下输出的特征图都需要送入Decoder，底层特征图在经过一次3x3的卷积和上采样后得到10x3的32通道特征，并进行如下操作：

（1）与进行concat。

（2）送入注意力模块。

（3）进行两次3x3卷积并进行上采样，得到。

即：

以上步骤仅适用于Decoder中所有中间层，以此类推，就可以得到不同尺度下的融合特征。最上层的处理应如图所示，只会进行一次3x3的卷积，并将得到的特征图进行一次20x1的卷积得到1x5的一通道的特征图，后接全连接层输出1\*8的经过one-hot编码的分类向量，代表输入信号所对应的人。