

Architect:

```
fusion_model(  
    (model_feat): model_features(  
        (L1): Linear(in_features=9, out_features=6, bias=True)  
        (L2): Linear(in_features=6, out_features=3, bias=True)  
    )  
    (model_img): model_image(  
        (L1): Conv2d(3, 64, kernel_size=(16, 16), stride=(8, 8))  
        (L2): Conv2d(64, 64, kernel_size=(4, 4), stride=(2, 2))  
        (L22): Conv2d(64, 64, kernel_size=(4, 4), stride=(2, 2))  
        (L3): Linear(in_features=1600, out_features=30, bias=True)  
    )  
    (decoder): decoder(  
        (L1): Linear(in_features=33, out_features=16, bias=True)  
        (L2): Linear(in_features=16, out_features=8, bias=True)  
        (L3): Linear(in_features=8, out_features=1, bias=True)  
    )  
)
```