Training and validation loss for 512×128 model