Training and validation loss for 512×256 model