1. The simpler models (fewer conv layers) seem to be behaving better – why?