

Why Dropout强大的大部分是由于施加到隐藏单元的掩码噪声，了解这一事实是重要的。

这可以看作是对输入内容的信息高度智能化、自适应破坏的一种形式，而不是对输入原始值的破坏。例如，如果模型学得通过鼻检测脸的隐藏单元 h_i ，那么丢失 h_i 对应于擦除图像中有鼻子的信息。模型必须学习另一种 h_i ，要么是鼻子存在的冗余编码，要么是脸部的另一特征，如嘴。传统的噪声注入技术，在输入端加非结构化的噪声不能够随机地从脸部图像中抹去关于鼻子的信息，除非噪声的幅度大到几乎能抹去图像中所有的信息。破坏提取的特征而不是原始值，让破坏过程充分利用该模型迄今获得的关于输入分布的所有知识。



