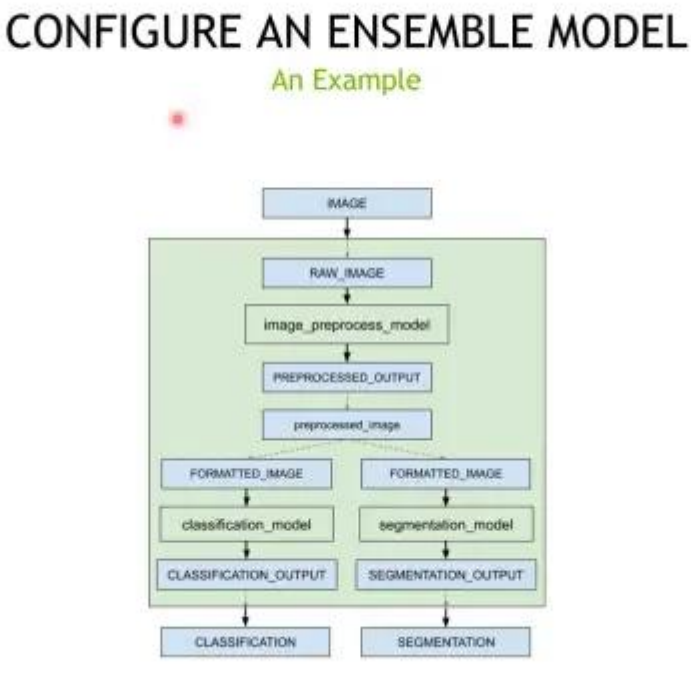
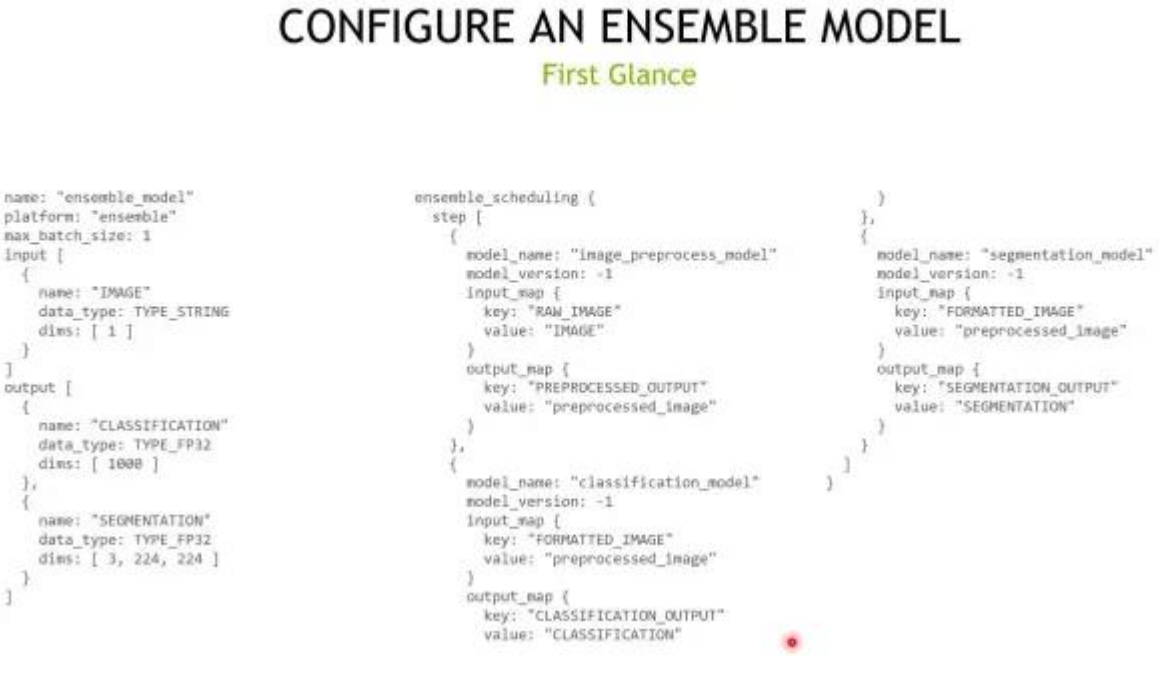
组合模型配置（ensemble model）

例子 1



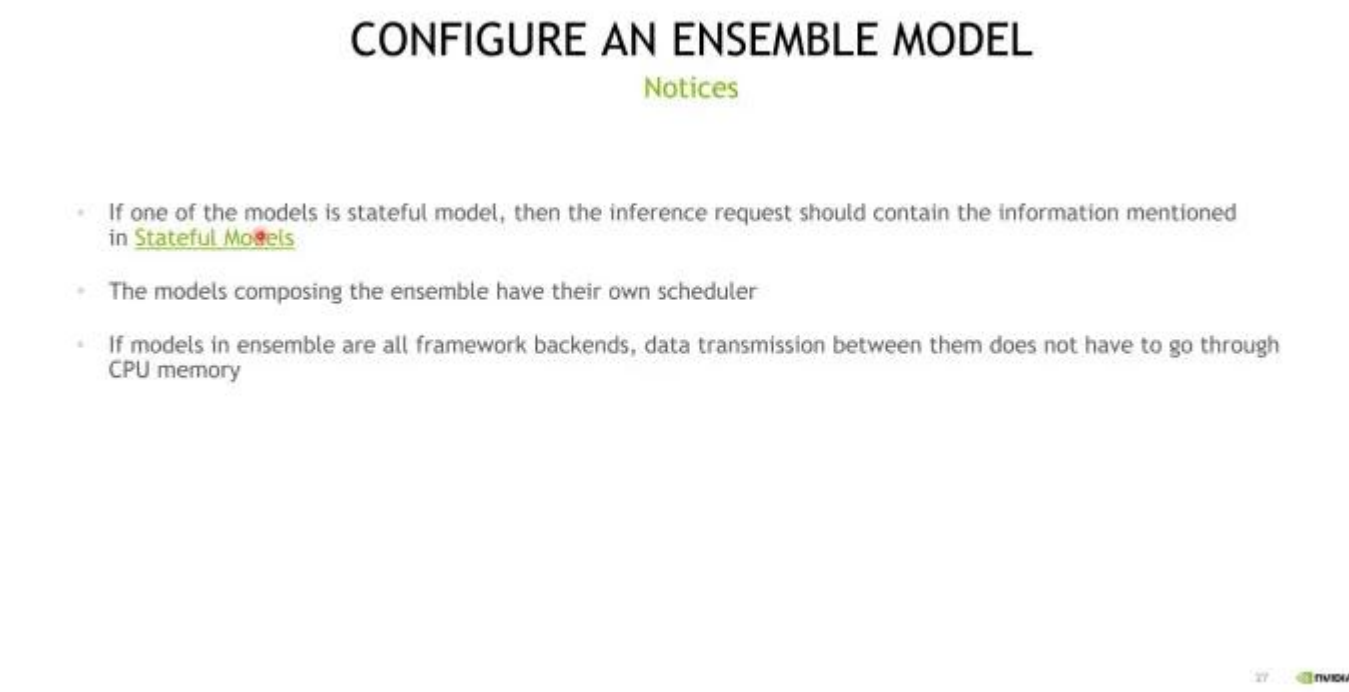
图片-预处理模型-分别进入不同的模型-输出不同的结果。



定义模型的输入输出，然后在 ensemble_scheduling 中定义不同的步骤，其中 step 中的 key 是本身的 input/output tensor 的名字；value 是 ensemble model 中的 Tensor 名字。

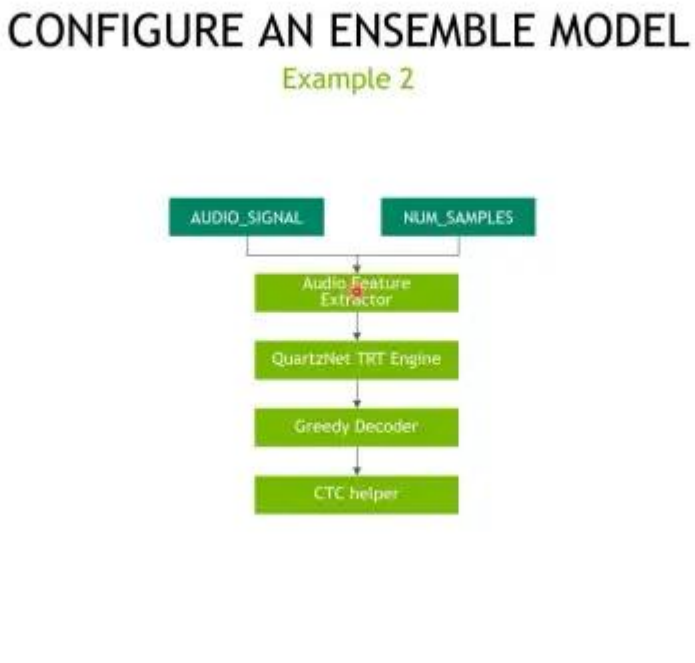
配置写完后，在 ensemble_model 的目录只能够新建一个版本目录，里面为空，然后放 config 文件。

注意事项：



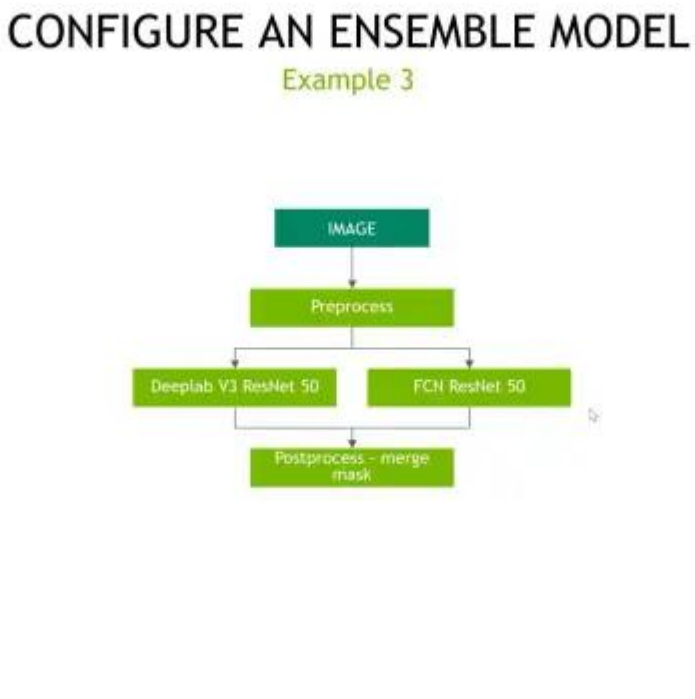
- 如果组合里有一个是 stateful 模型，那么整个 pipeline 都成为 stateful 模型，推理请求需要符合 stateful model 的规则。
- 每个子模块有各自的调度器。
- 如果每个子模块都是 framework backend，则传输使用 GPU 进行，否则可能通过 cpu 内存。

例子 2



语音识别流程。

例子 3



预处理-分支 1 分割模型-分支 2fcn 分割-合并拼接