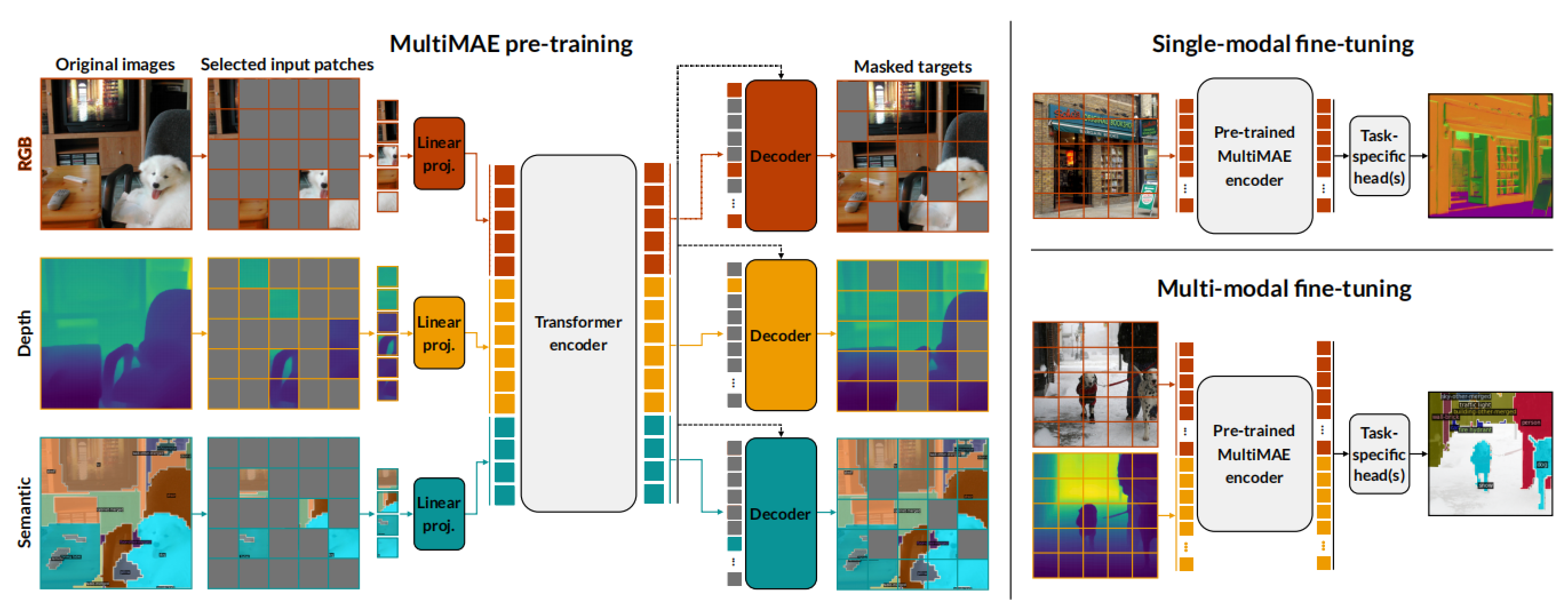
## 讨论内容：

1. 模态图像重建技术探讨

《MultiMAE: Multi-modal Multi-task Masked Autoencoders》



多模态图像重建，该文献提出了一个模型，通过线性映射到transformer encoder的输入端，进行位置编码和模态编码，提取关键特征，并分别分到三个decoder里进行重建。

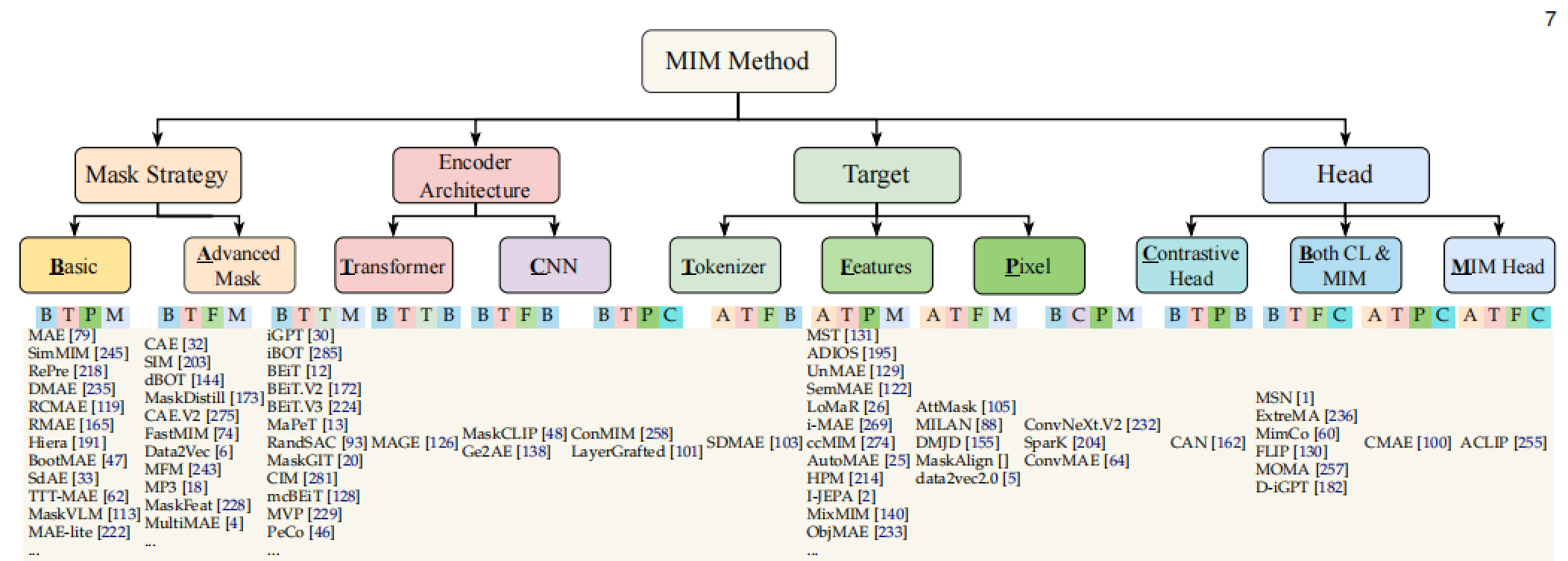
微调阶段：可进行单模态微调或多模态微调，即通过两种模态重建另一种模态图像。

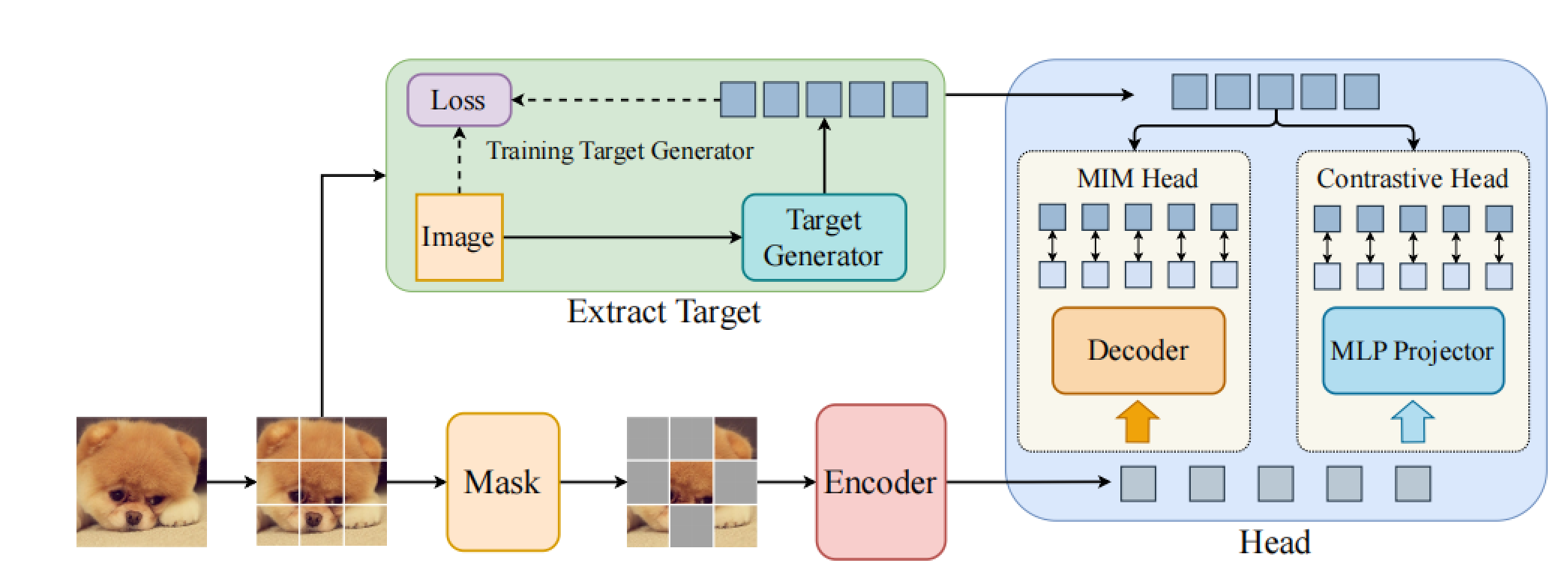
2、模态混合，missing modality

讨论了预训练过程和现实场景下的模态遗失问题，即如何在实际应用场景中处理模态遗失，例如在机器人应用中，预训练考虑了多种模态，实际应用时采集到的数据可能只有一个或两个模态可用。

思路：将预训练的encoder应用于单模态或多模态任务，使其能够自适应推理。

3、MIM综述文献





## 待办：

1. 将思路绘制成框架图，上传至github；
2. 确定思路后开始着手复现，修改代码。