

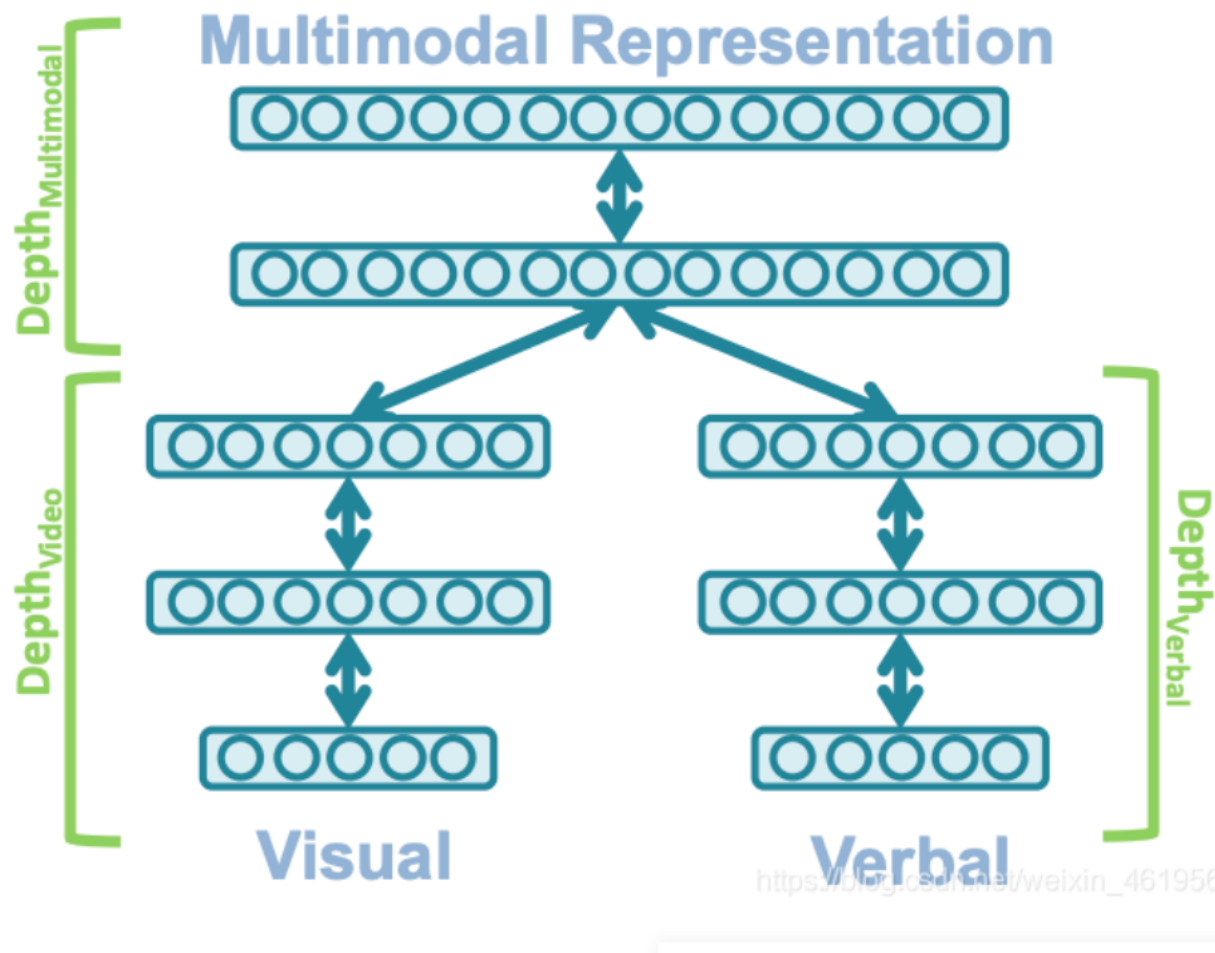
Multimodal Learning

Sinkoo

March 28, 2022

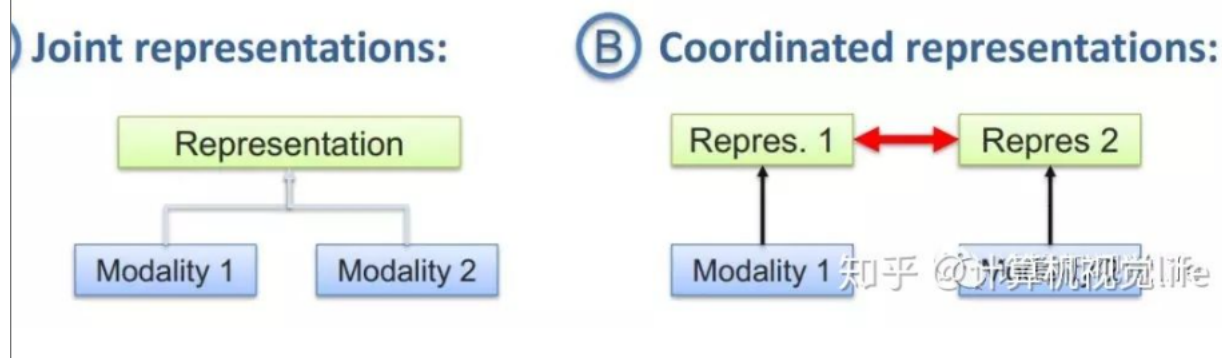
1 MultiModal presentation

学习将多模态数据整合到一个特征表示：互补，减少冗余：



例子是视觉与听觉。

主要有两大研究方向：联合表示（映射到一个统一的空间），协同学习（映射到的空间，但是两个空间相互约束）



2 Translation

一个模态的信息转化为另一个：主要有：机器翻译，唇语和语音翻译；图片描述，视频描述；语音合成等。

3 Alignment

多模态的对齐对来自同一个实例的不同模态信息寻找对应关系。

4 Co-learning

协同学习是指使用一个模态的信息来辅助另一模态进行学习。

协同训练：用一个模态对数据集标记进行扩充给另一个模态学习。：在机器学习中有结合两个模态进行半监督学习：给两个模态分别训练学习器，一个模态识别出的有把握的样本进行标记，并把标记的样本给另一个模态学习器训练。迭代互补以提高性能。