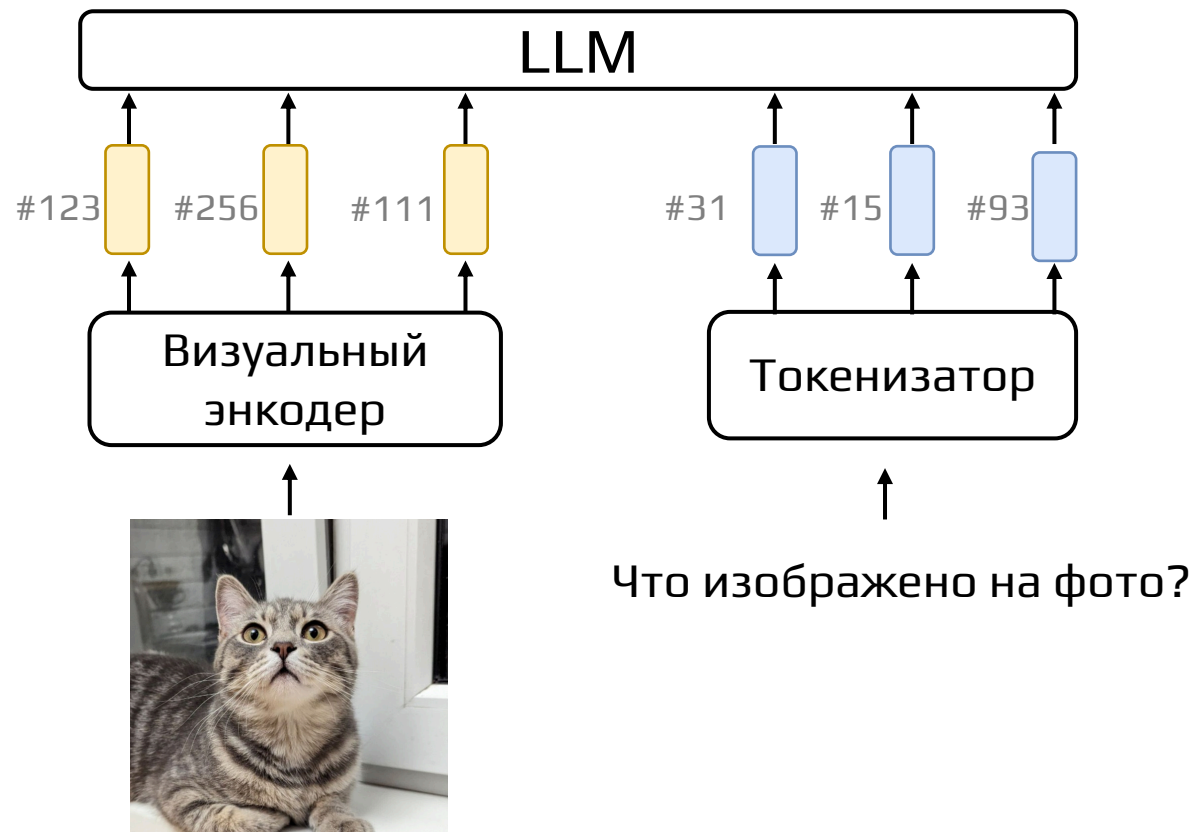


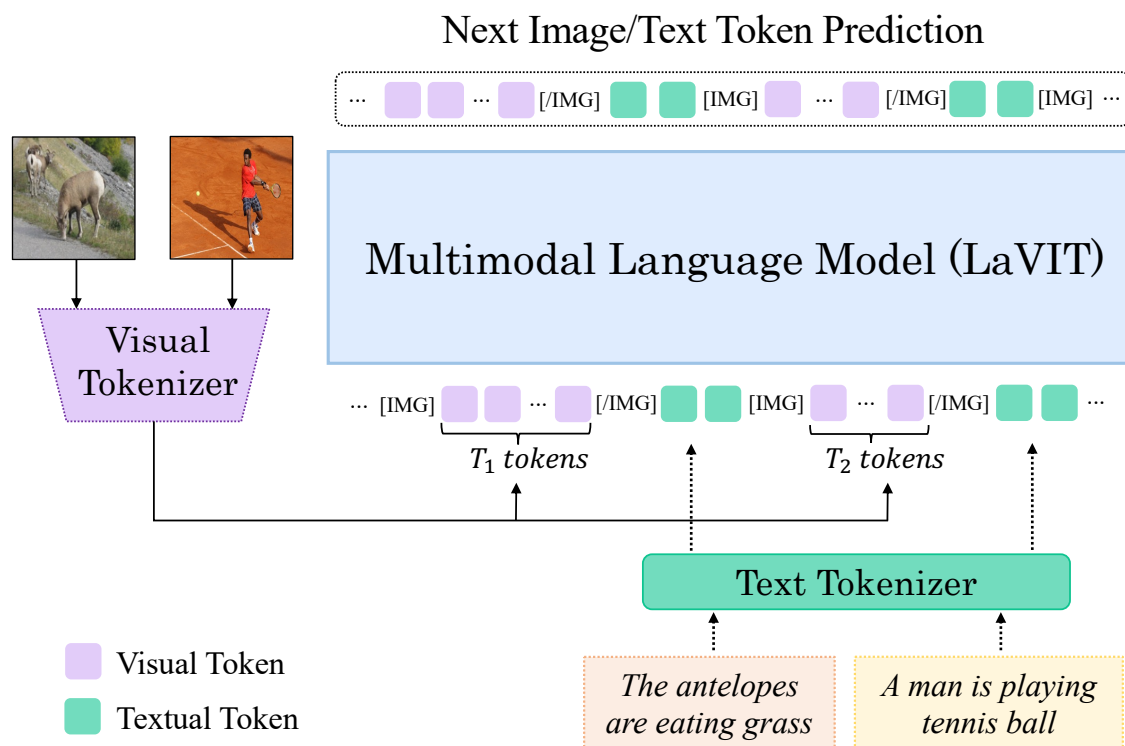
Курс “Мультимодальные модели”

Предпосылки

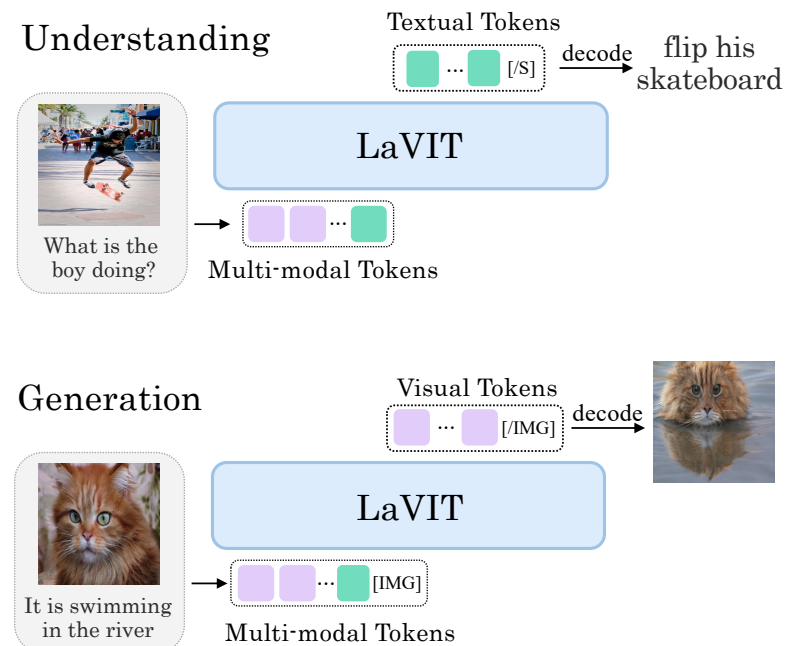
- Ранее рассмотрели модели, в которых визуальные эмбединги подаются в VLM как векторы
- А можно ли заранее задать некоторый набор визуальных токенов и использовать только их?



LaVIT (2024).

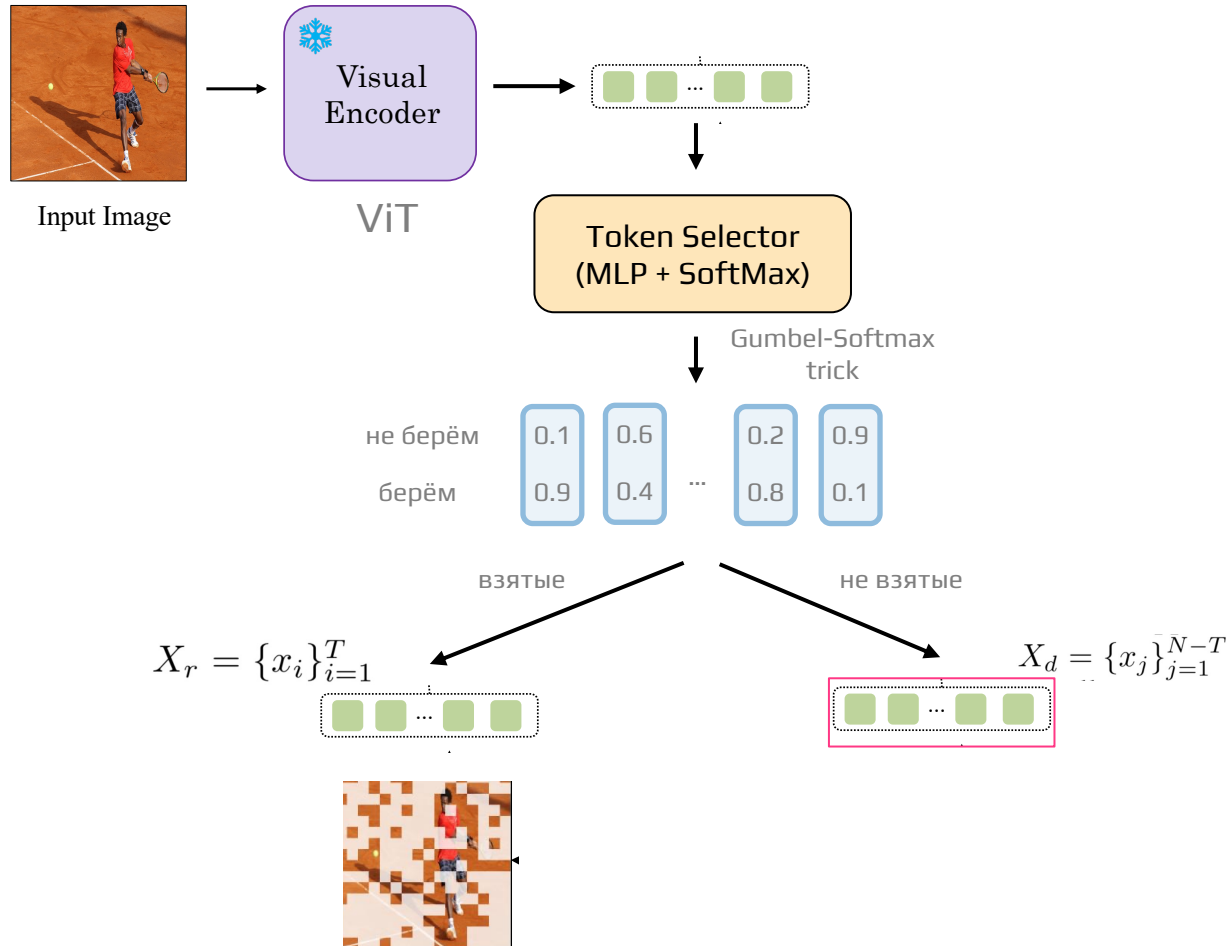


(a) Pre-training

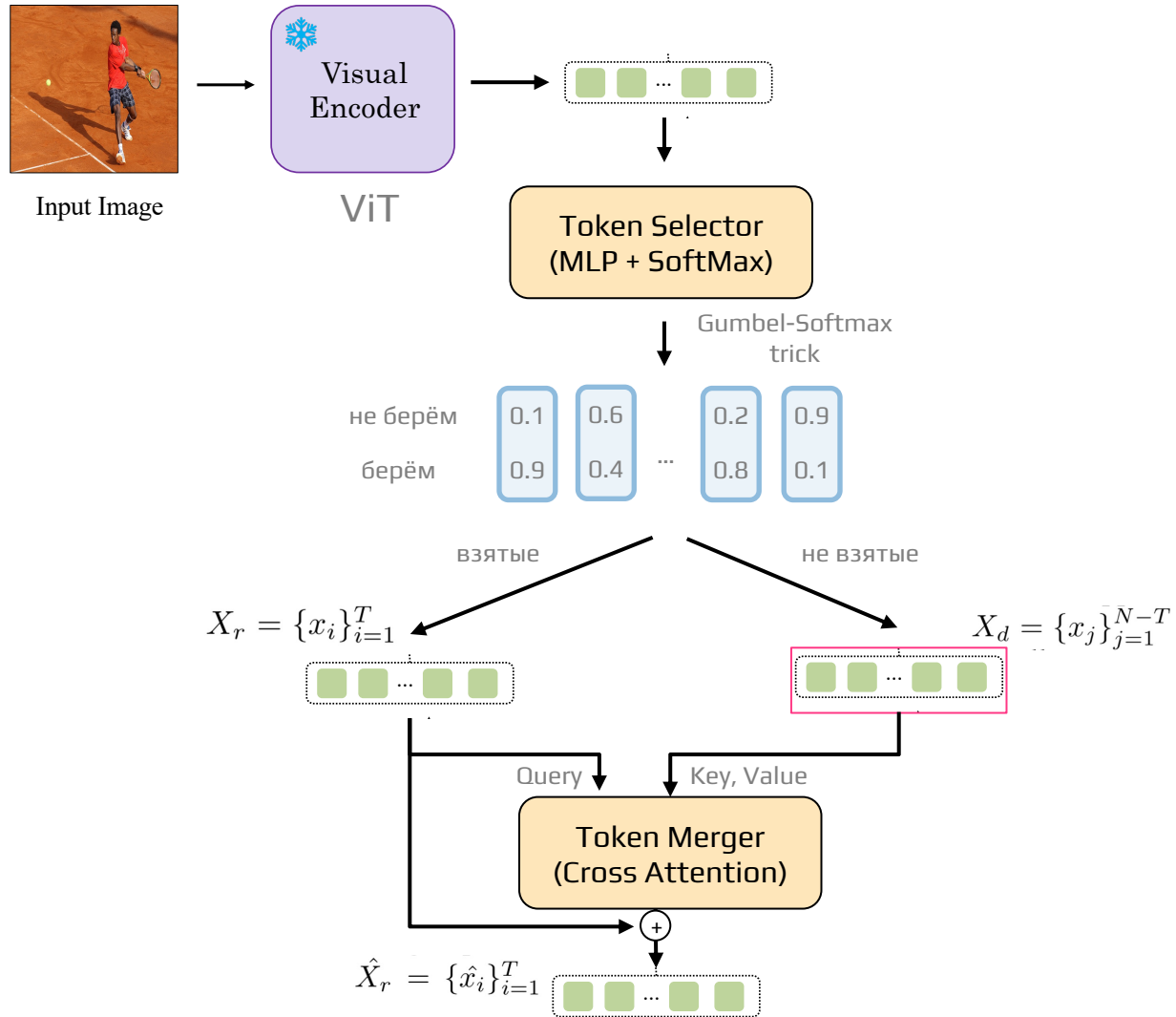


(b) Inference

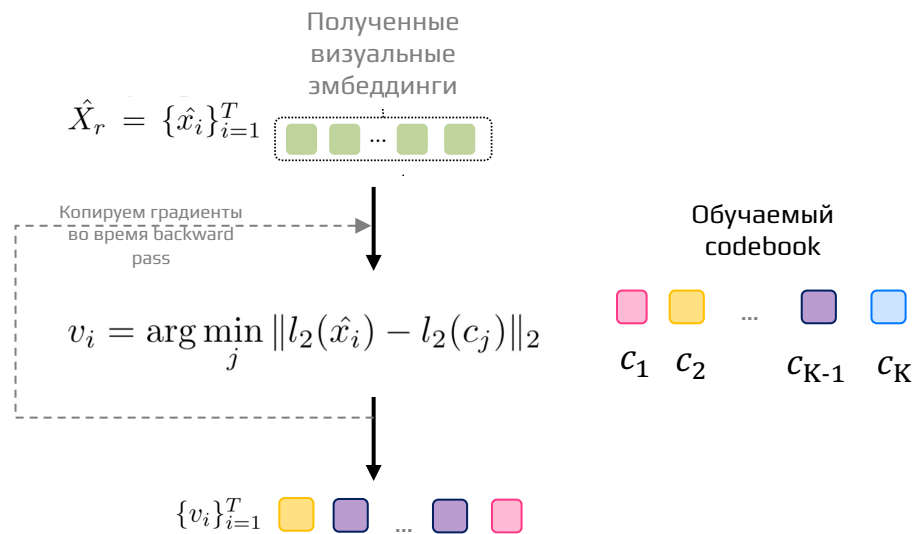
LaVIT (2024). Token Selector



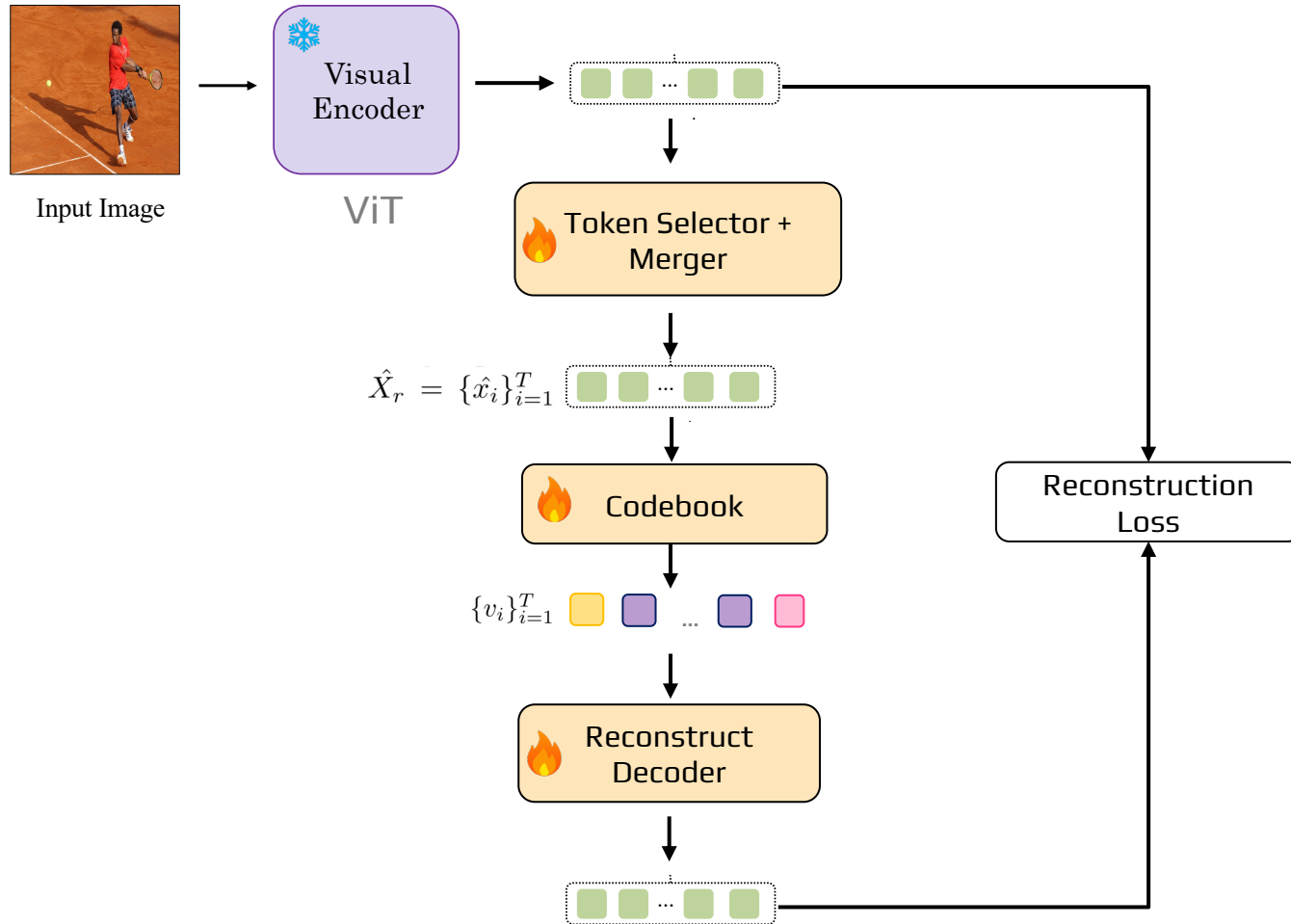
LaVIT (2024). Token Merger



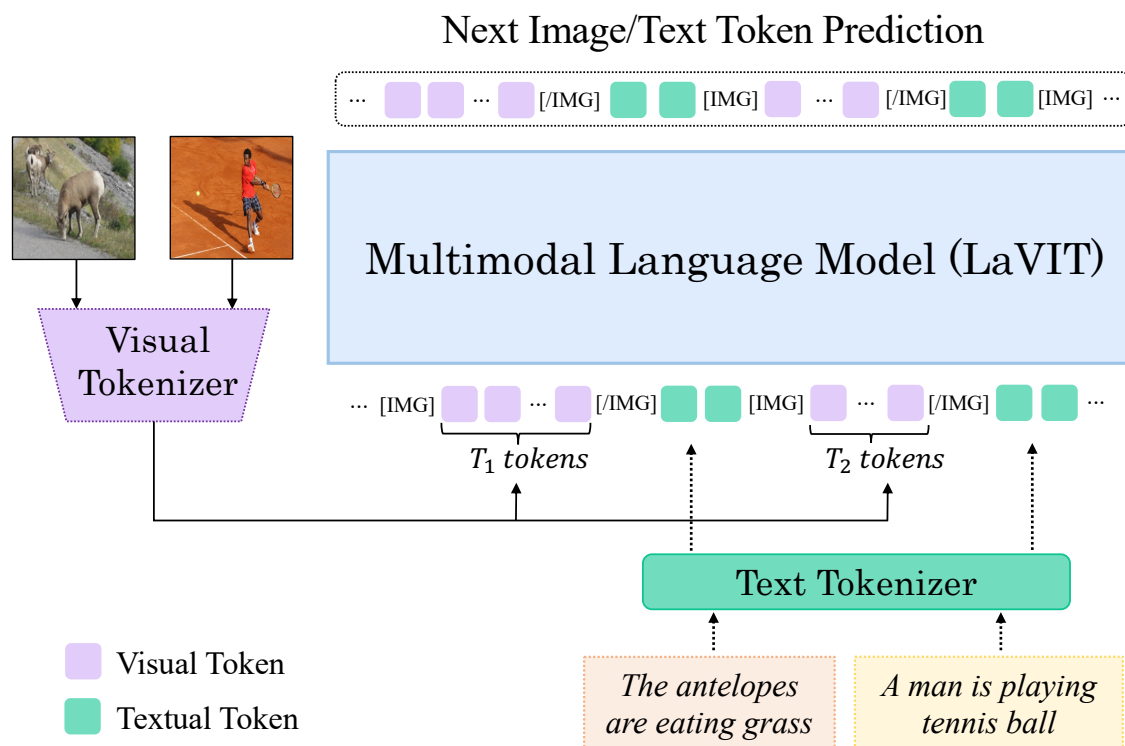
LaVIT (2024). Дискретизация векторов



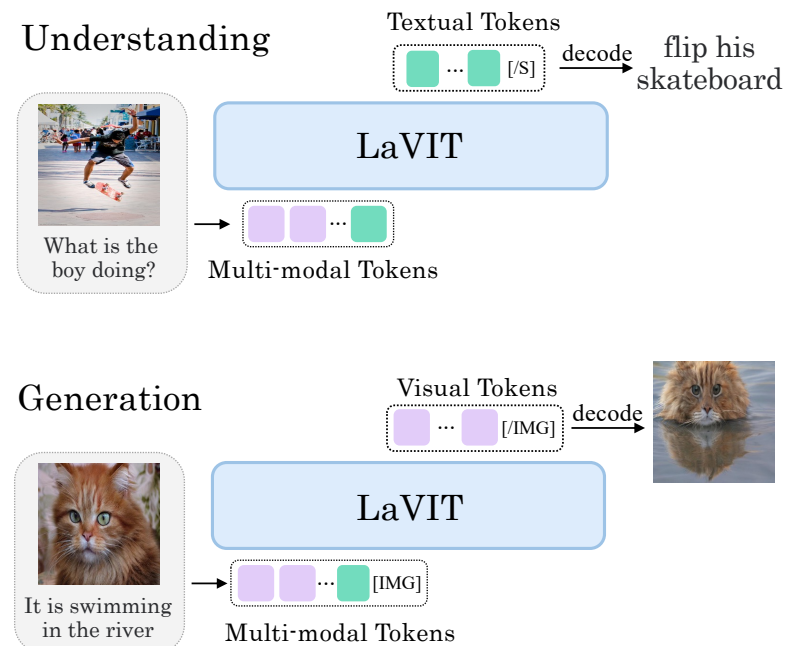
LaVIT (2024). 1-й этап обучения



LaVIT (2024). 2-й этап обучения



(a) Pre-training



(b) Inference

LaVIT (2024). Восстановление изображений

