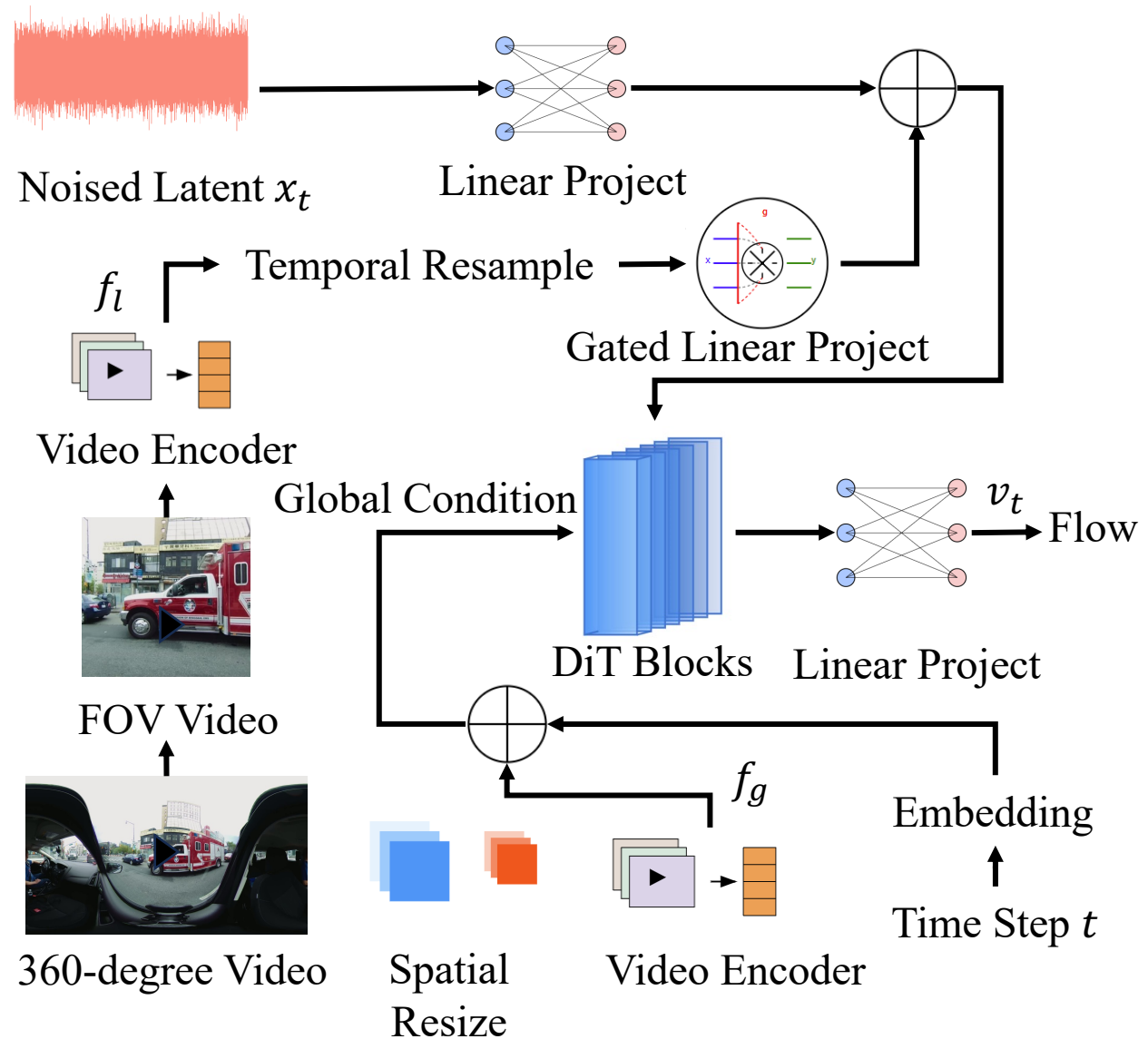


(a) Self-Supervised Coarse-to-Fine Pre-training



(b) 360-degree Video-Guided Fine-tuning