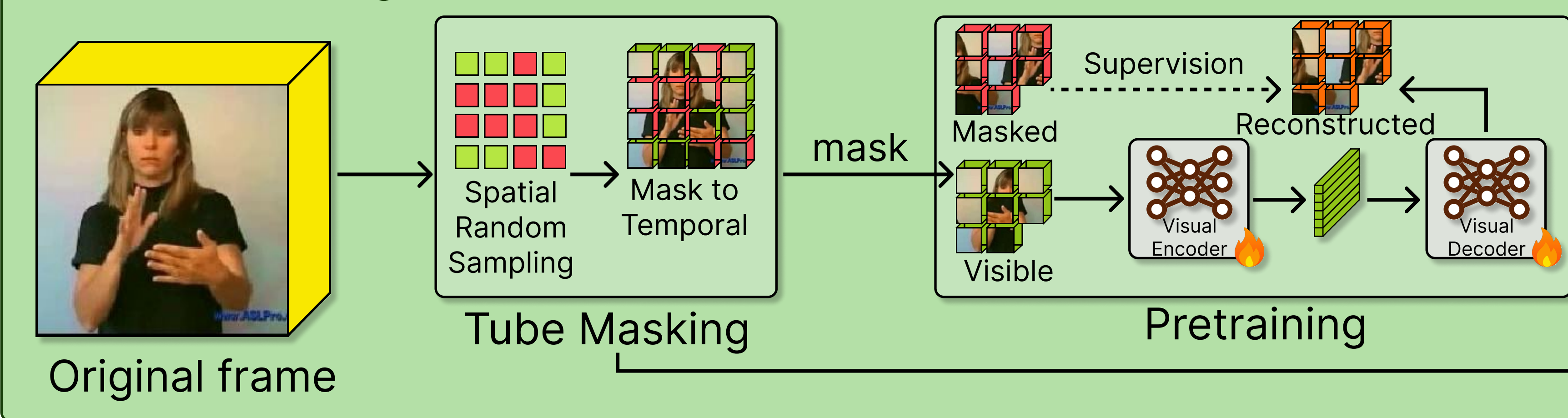
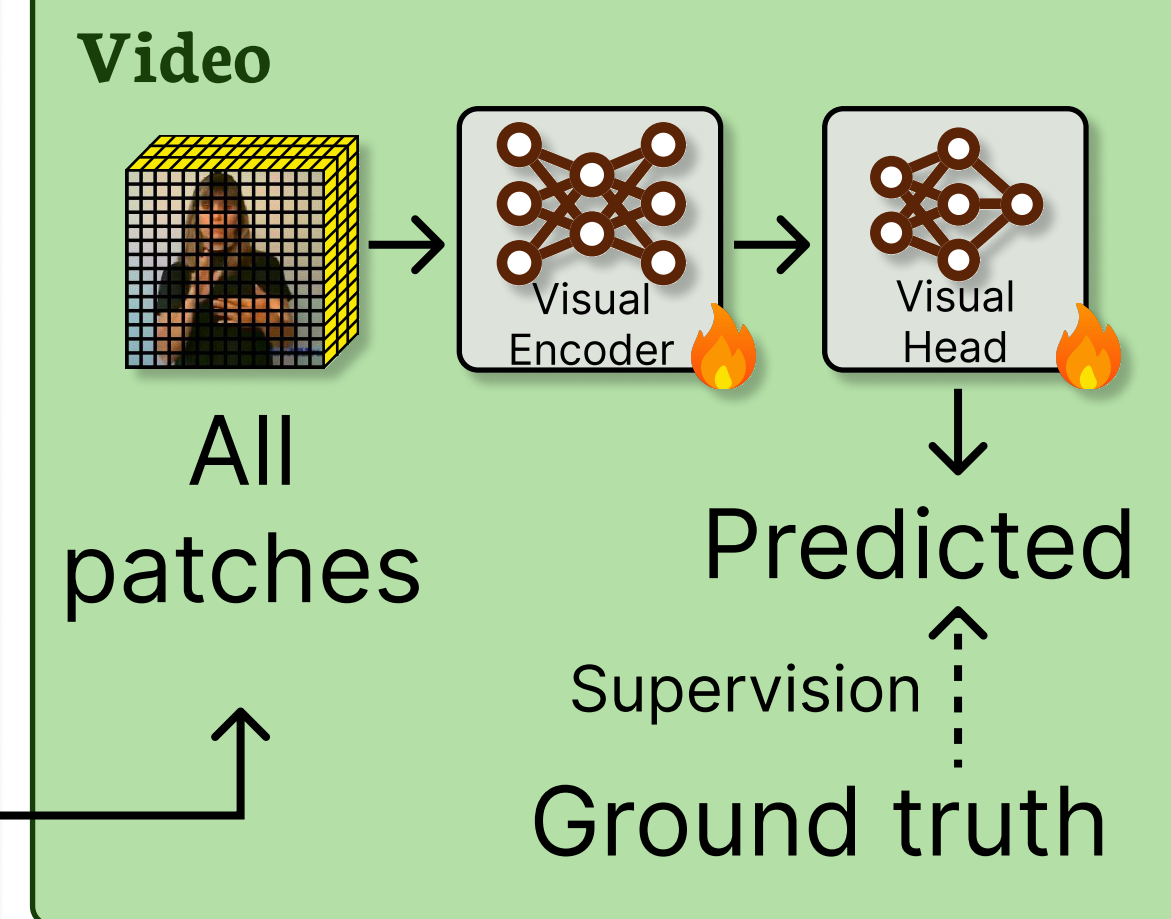


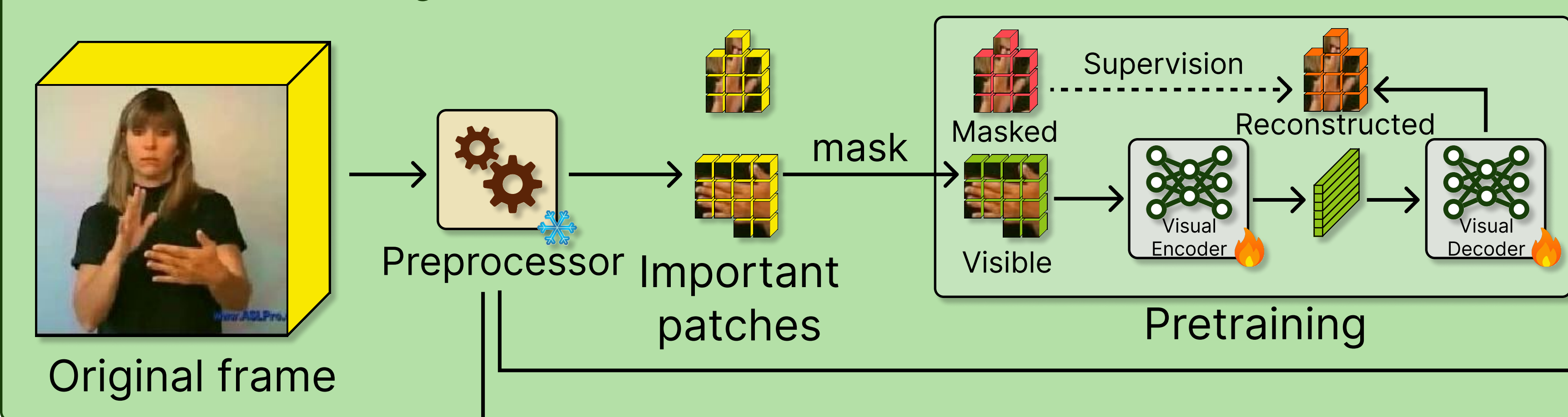
### Tube Mask Pretraining - Video



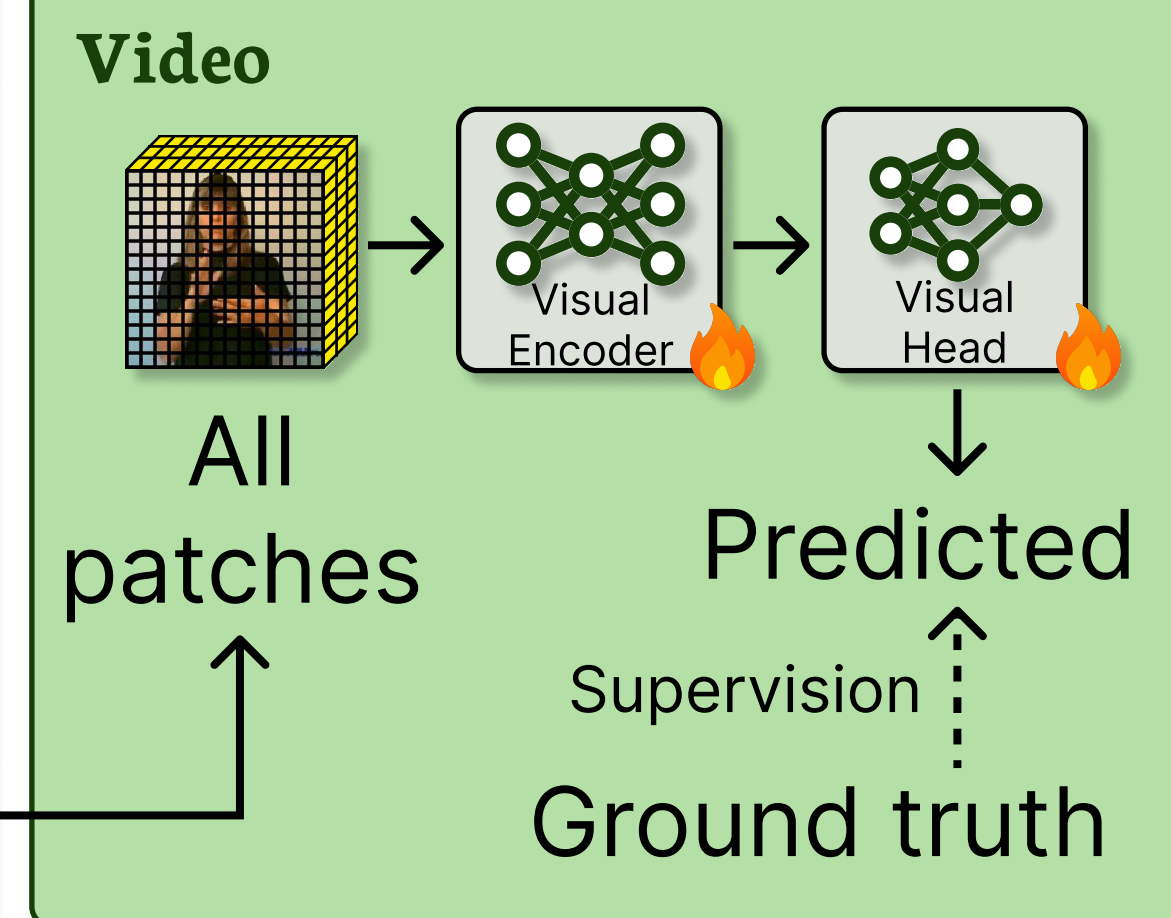
### Uni-modal Fine-tuning - Video



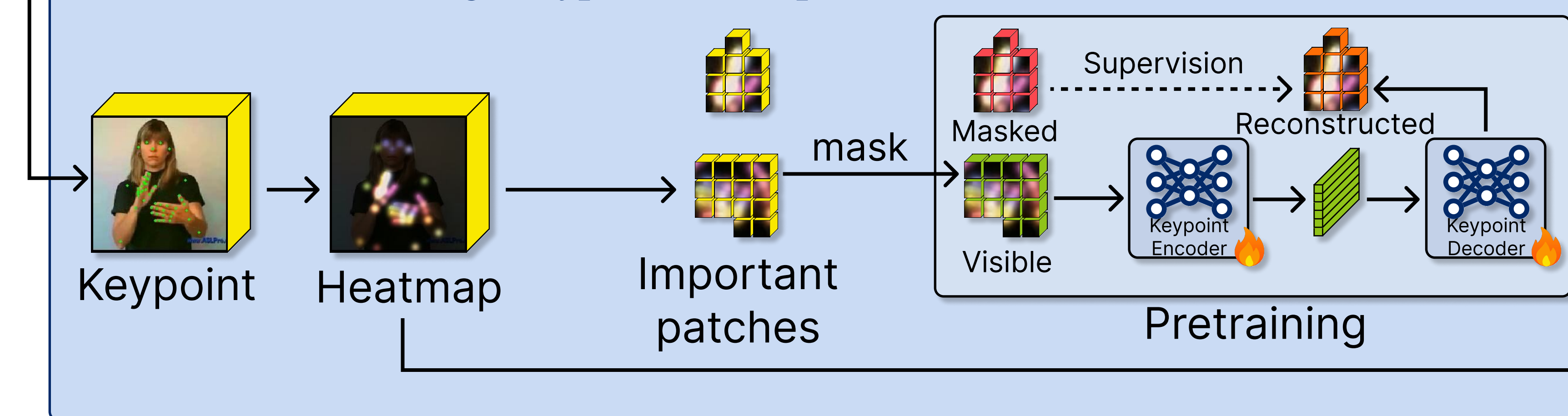
### Hand-Guided Pretraining - Video



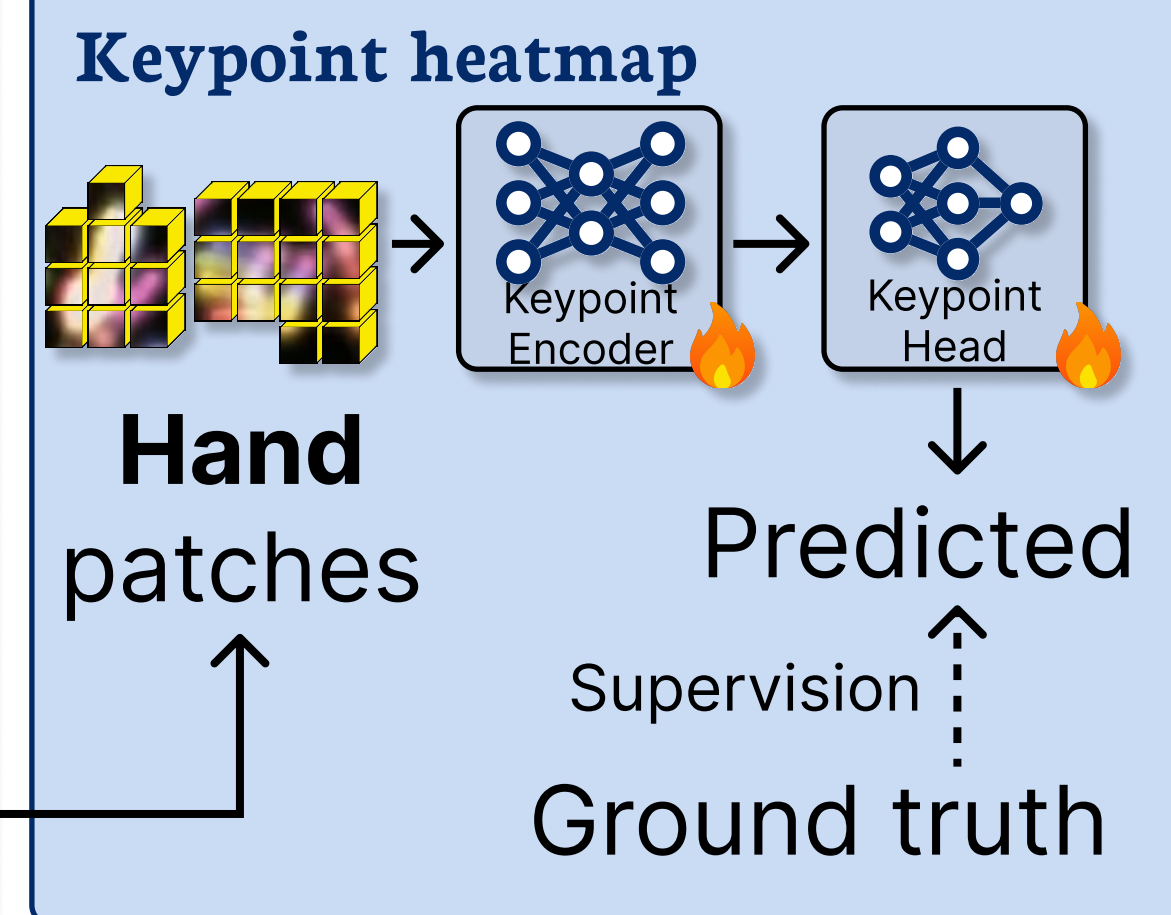
### Uni-modal Fine-tuning - Video



### Hand-Guided Pretraining - Keypoint heatmap



### Uni-modal Fine-tuning - Keypoint heatmap



### Multi-modal Fusion

