

Condition Injection Methods

Part 5/7: Architecture



Time Embedding

Sinusoidal positional encoding + MLP



Class Embedding

Learned embedding for each class



Text Conditioning

CLIP or T5 text encoder → cross-attention



Adaptive Group Norm

Scale and shift from condition embedding



FiLM Layers

Feature-wise Linear Modulation



Concatenation

Directly concat condition to input channels



Cross-Attention

Text tokens attend to image features



Hybrid Approaches

Combine multiple conditioning methods for optimal performance

Categorization by Type



Temporal

Time Embedding



Categorical

Class Embedding



Textual

Text + Cross-Attention



Modern models often combine multiple injection methods for best results