Visualization and Analysis of Learned Flow Gates

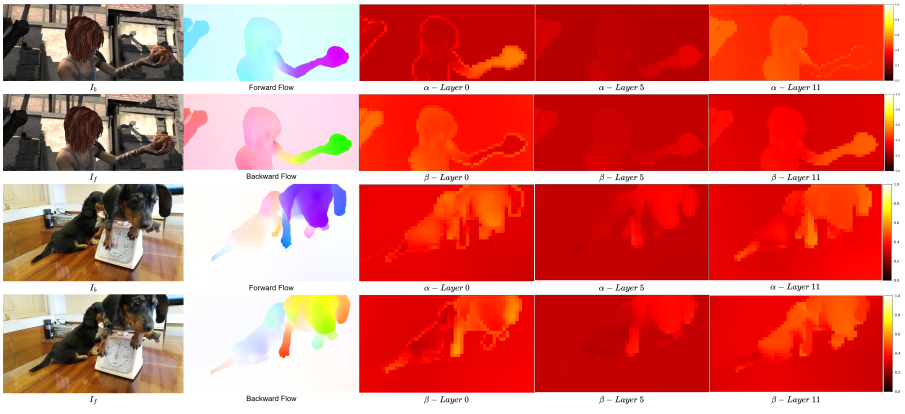

Figure 1: Visualization of the learned forward (α) and backward (β) flow gates at Layers 0, 5, and 11 of Flow-DiT.

To provide further insight into the proposed flow-guided gating mechanism, we visualize the learned forward and backward flow gates α and β at different depths of the Flow-DiT transformer. As defined in Section 3.4, α and β are obtained by applying a sigmoid function to the forward and backward flow tokens, transformed using learned linear weight matrices W_α and W_β, respectively, and therefore take values in the range $[0, 1]$. Higher gate values indicate that the corresponding flow tokens have a stronger contribution during temporal attention.

The visualizations in Fig. 1 show that the learned gates are not spatially uniform. Instead, both α and β exhibit localized responses that follow regions with stronger motion or more challenging temporal correspondence. This behavior suggests that the model does not uniformly inject the estimated flow into all spatial tokens, but selectively modulates its contribution according to the local motion context. The gate activations also evolve across the network depth. For example, the spatial patterns observed at Layer 0 differ from those at Layers 5 and 11, indicating that flow information is utilized differently at different stages of the diffusion transformer.