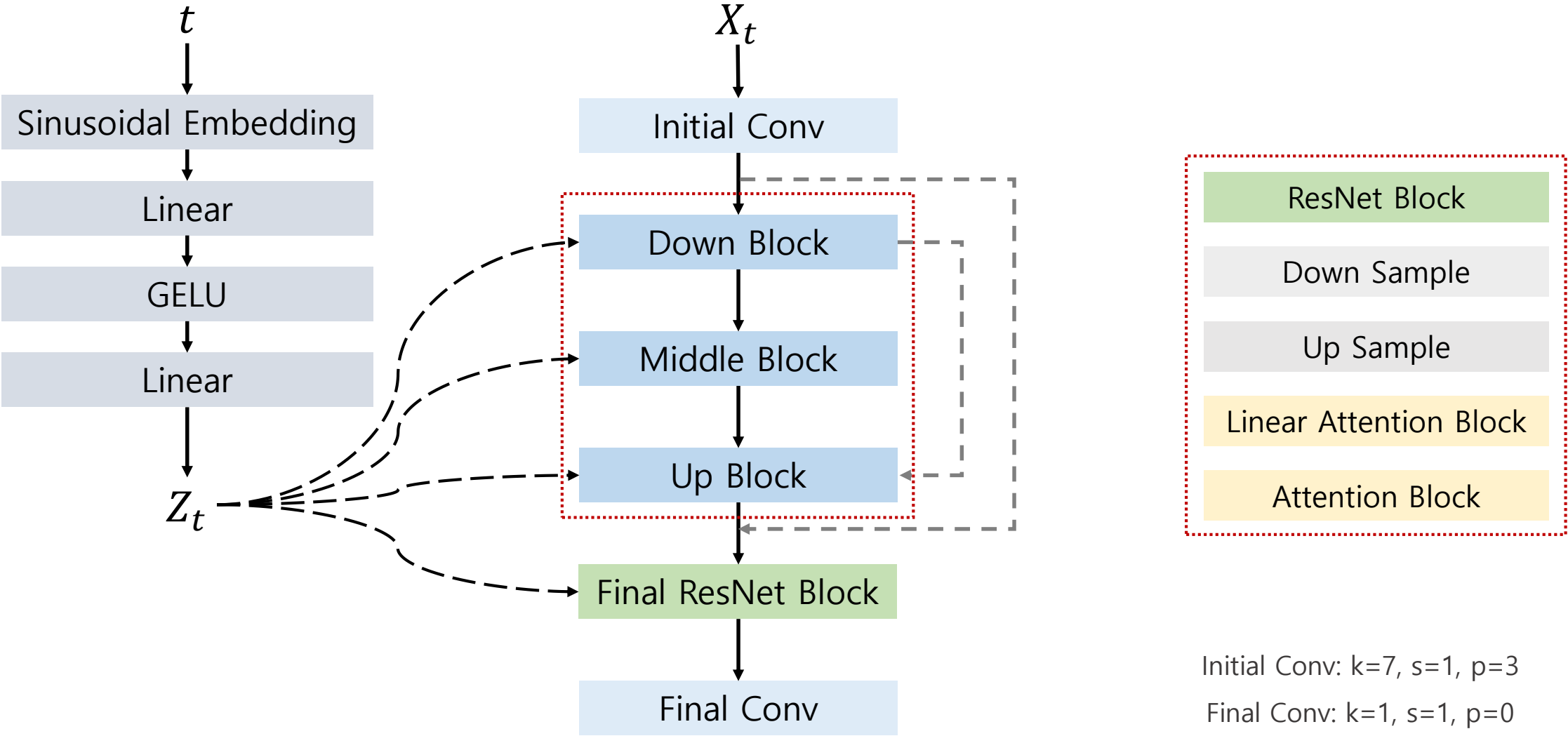


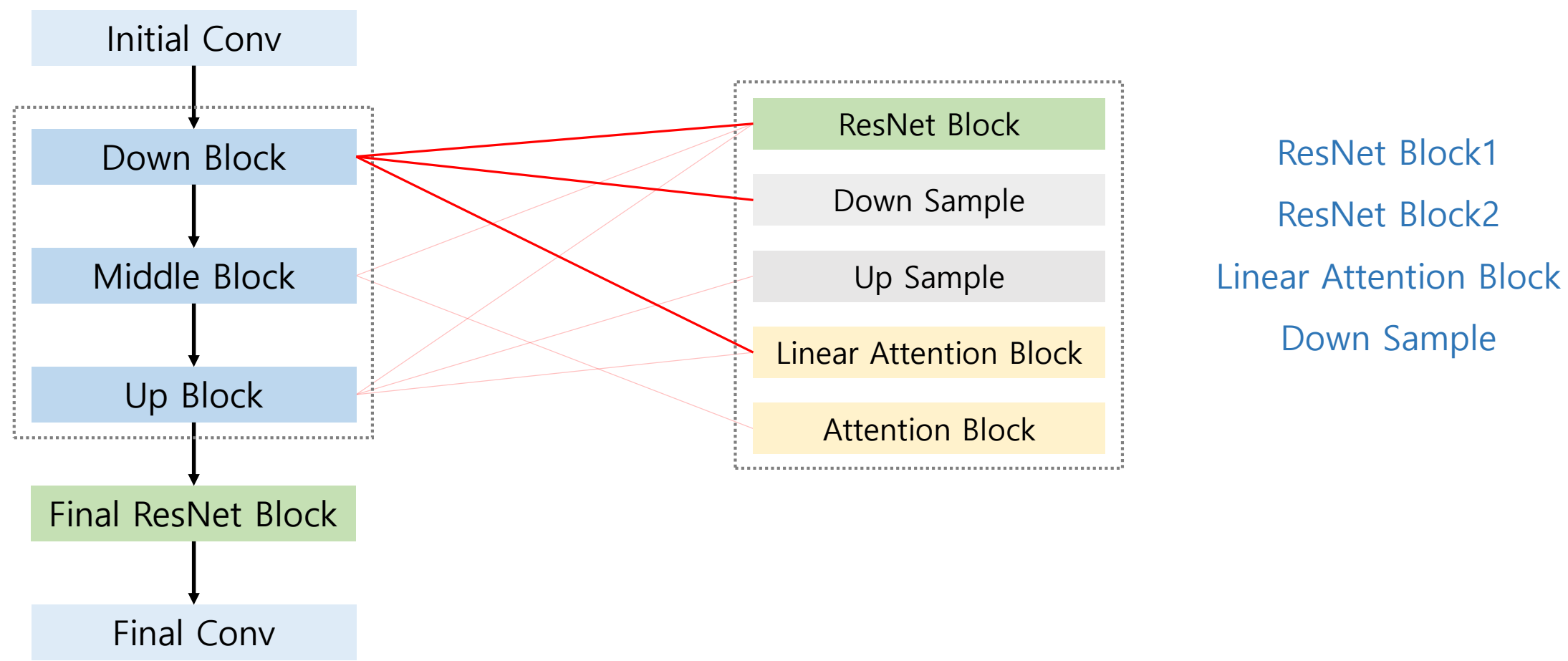
U-Net Structure of DDPM

<https://github.com/lucidrains/denoising-diffusion-pytorch>

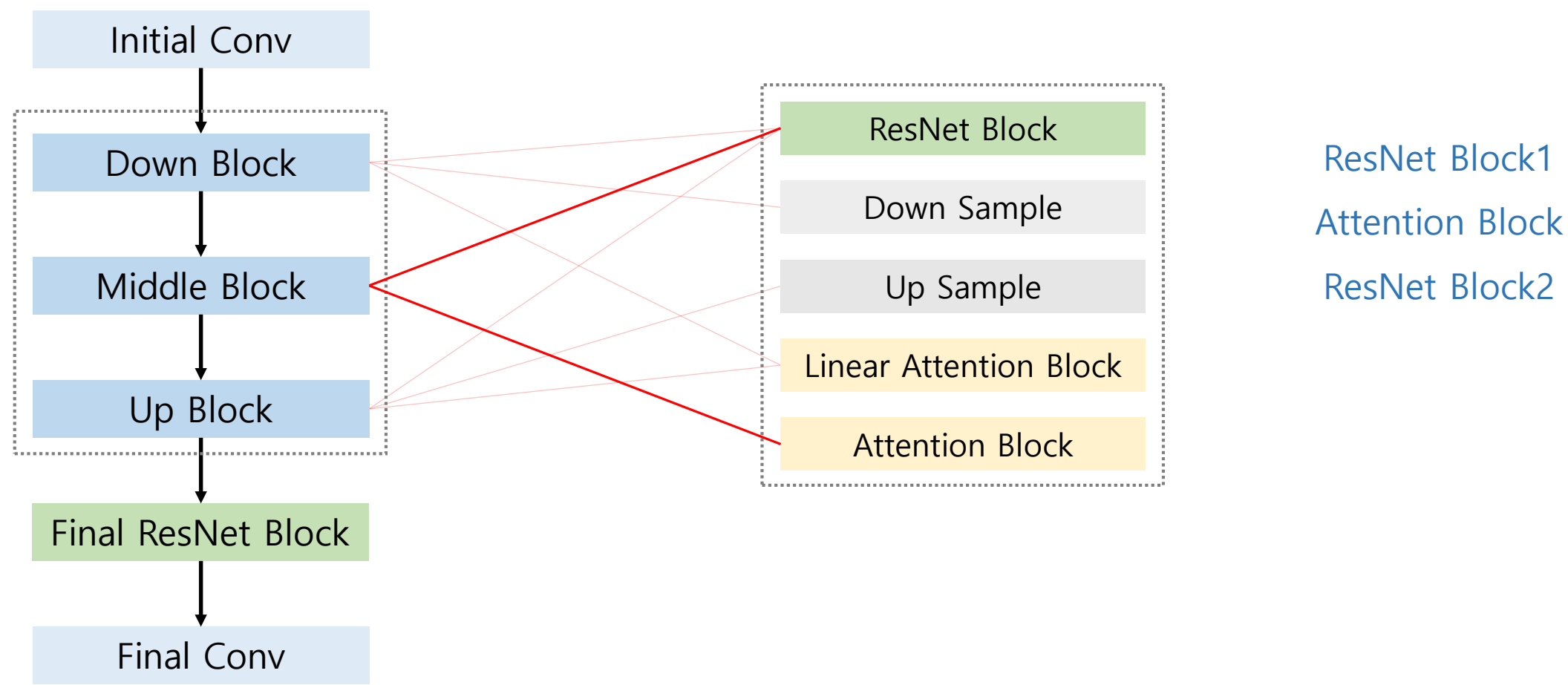
Overall Structure



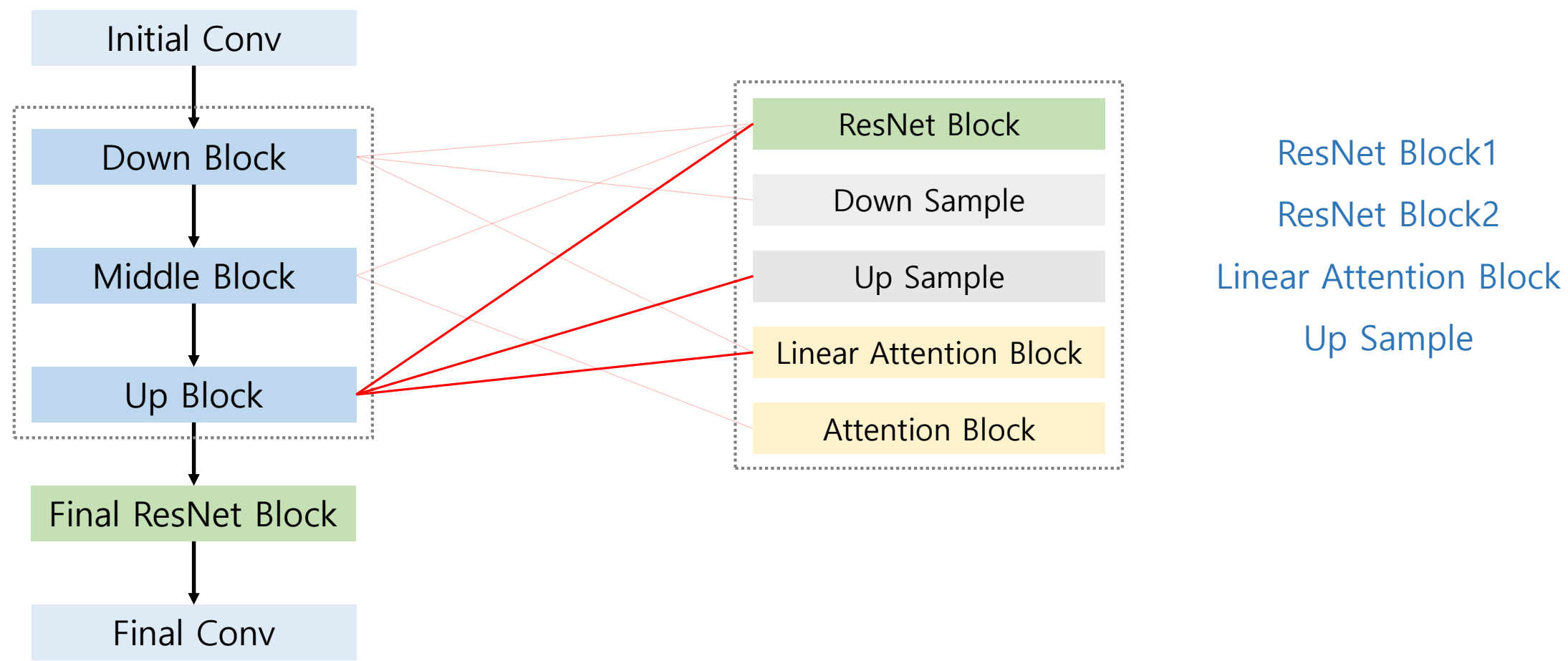
Overall Structure



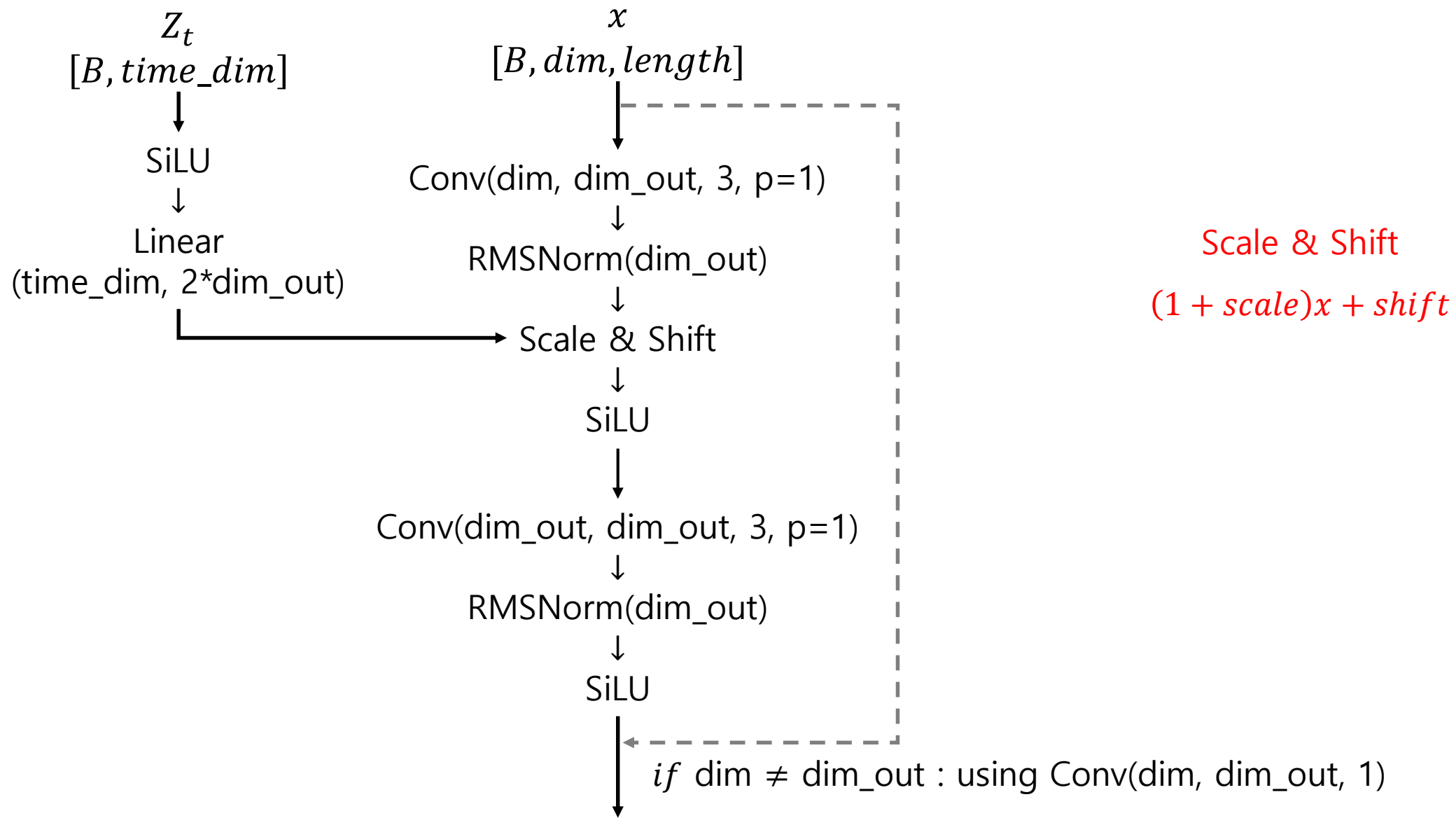
Overall Structure



Overall Structure



ResNet Block (dim, dim_out)



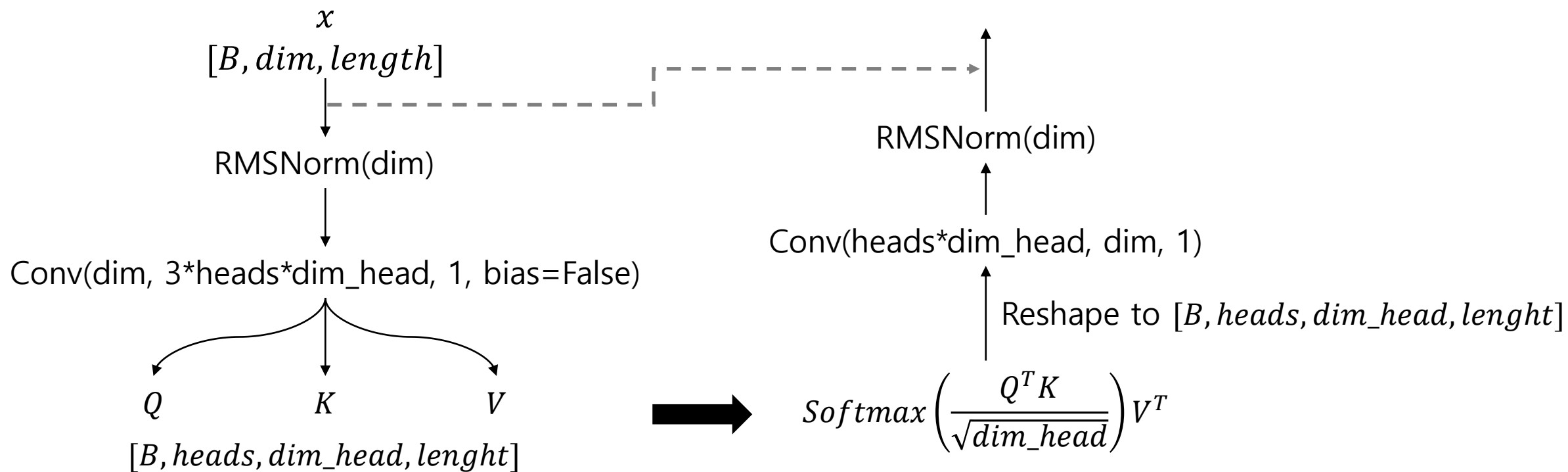
Down Sample (dim, dim_out)

$$\overset{x}{[B, dim, length]} \longrightarrow \text{Conv}(dim, dim_out, 4, s=2, p=1)$$

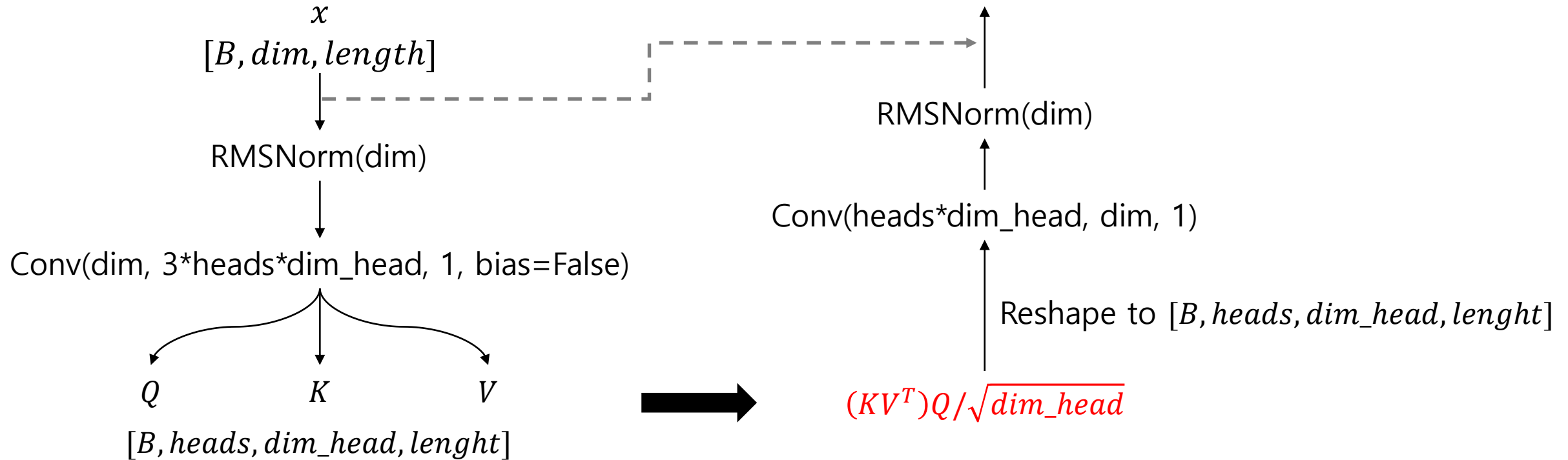
Up Sample (dim, dim_out)

$$\overset{x}{[B, dim, length]} \longrightarrow \text{Upsample}(scale_factor=2, mode='nearest') \longrightarrow \text{Conv}(dim, dim_out, 3, p=1)$$

Attention Block (dim, heads, dim_head)

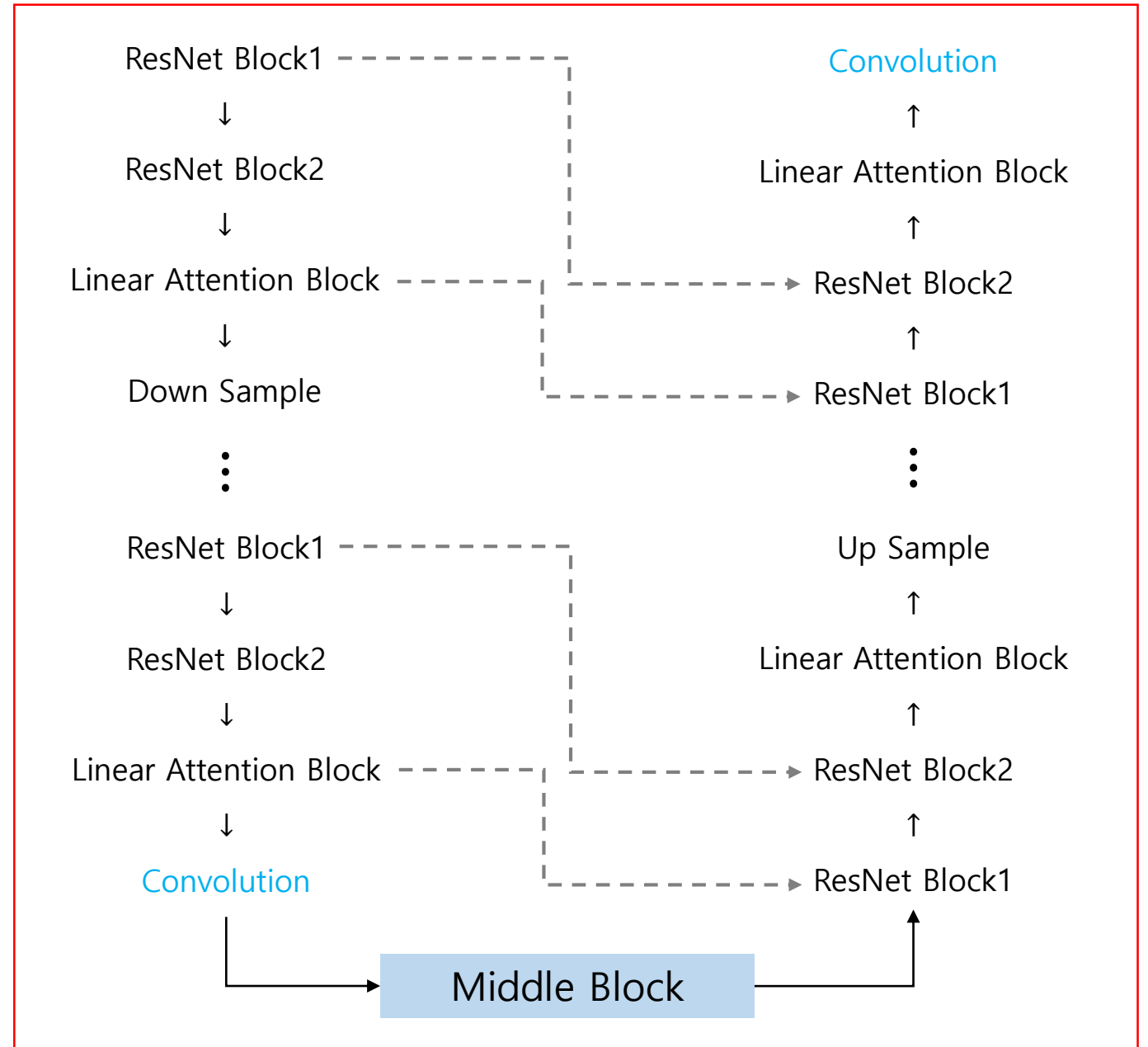
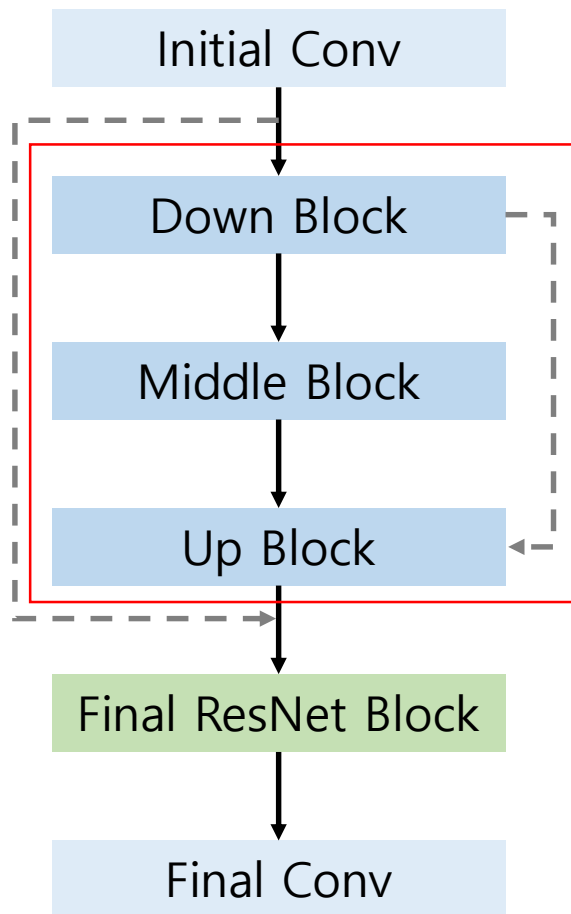


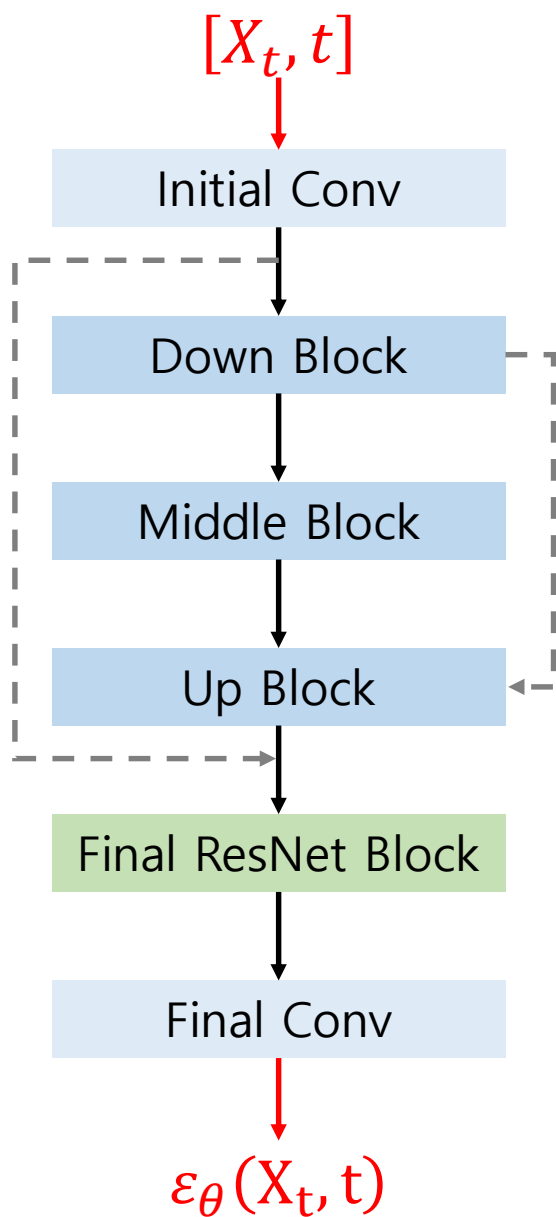
Linear Attention Block (dim, heads, dim_head)



$$Q = \text{Softmax}(Q, \text{dim} = -2)$$

$$K = \text{Softmax}(K, \text{dim} = -1)$$





$$q(X_t|X_{t-1}) = \mathcal{N}(X_t; \sqrt{1 - \beta_t}X_{t-1}, \beta_t I)$$

$$q(X_t|X_0) = \mathcal{N}(X_t; \sqrt{\bar{\alpha}_t}X_0, (1 - \bar{\alpha}_t)I)$$

$$X_t = \sqrt{\bar{\alpha}_t}X_0 + \sqrt{1 - \bar{\alpha}_t}\varepsilon_t$$

$$\varepsilon_\theta(X_t, t) \approx \varepsilon_t$$

$$D_{KL}(q(X_{t-1}|X_t, X_0) \parallel p_\theta(X_{t-1}|X_t))$$

$$\approx \|\varepsilon_t - \varepsilon_\theta(X_t, t)\|$$

Training Process

