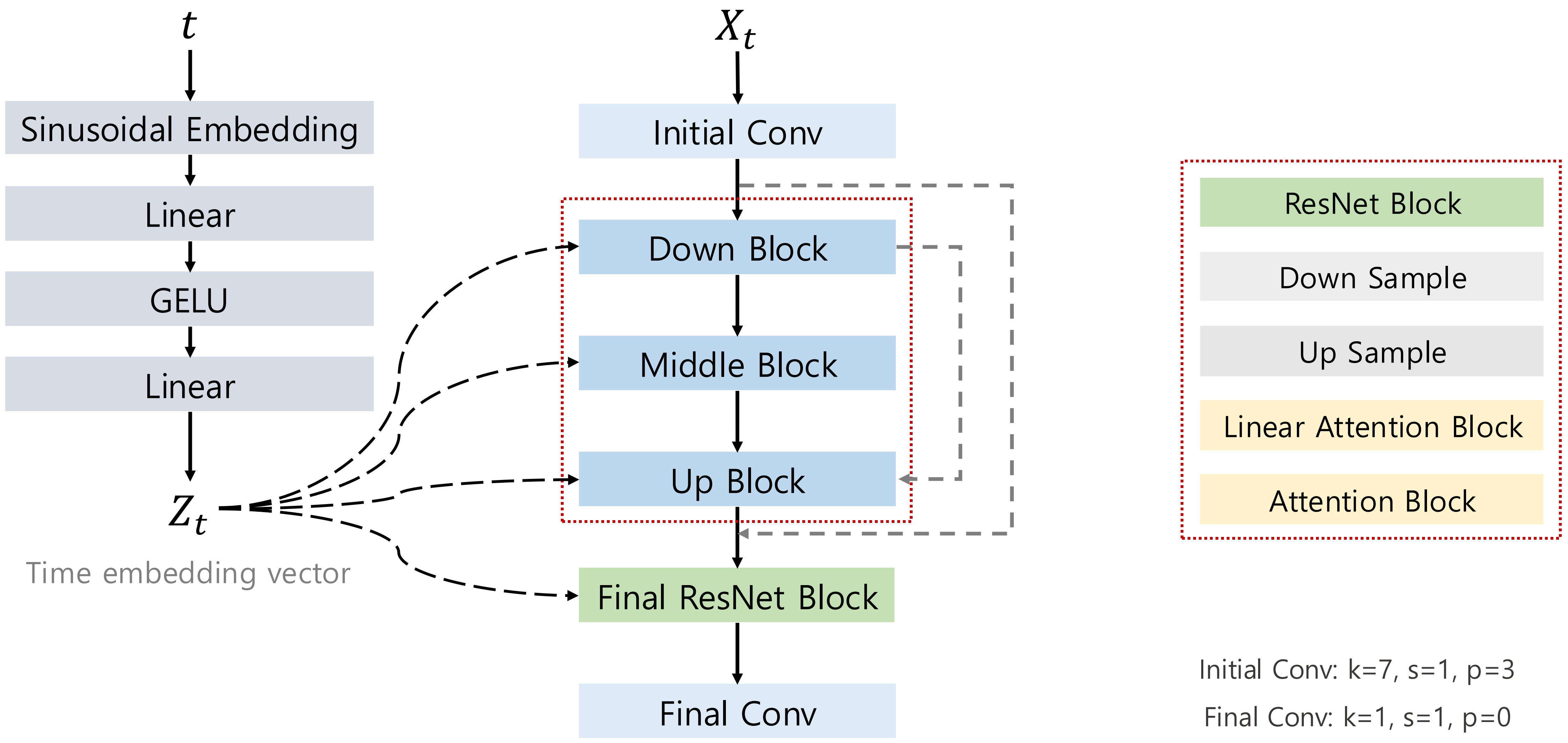
Denoising Diffusion Probabilistic Models

U-Net Structure of DDPM & Processes of DDPM

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

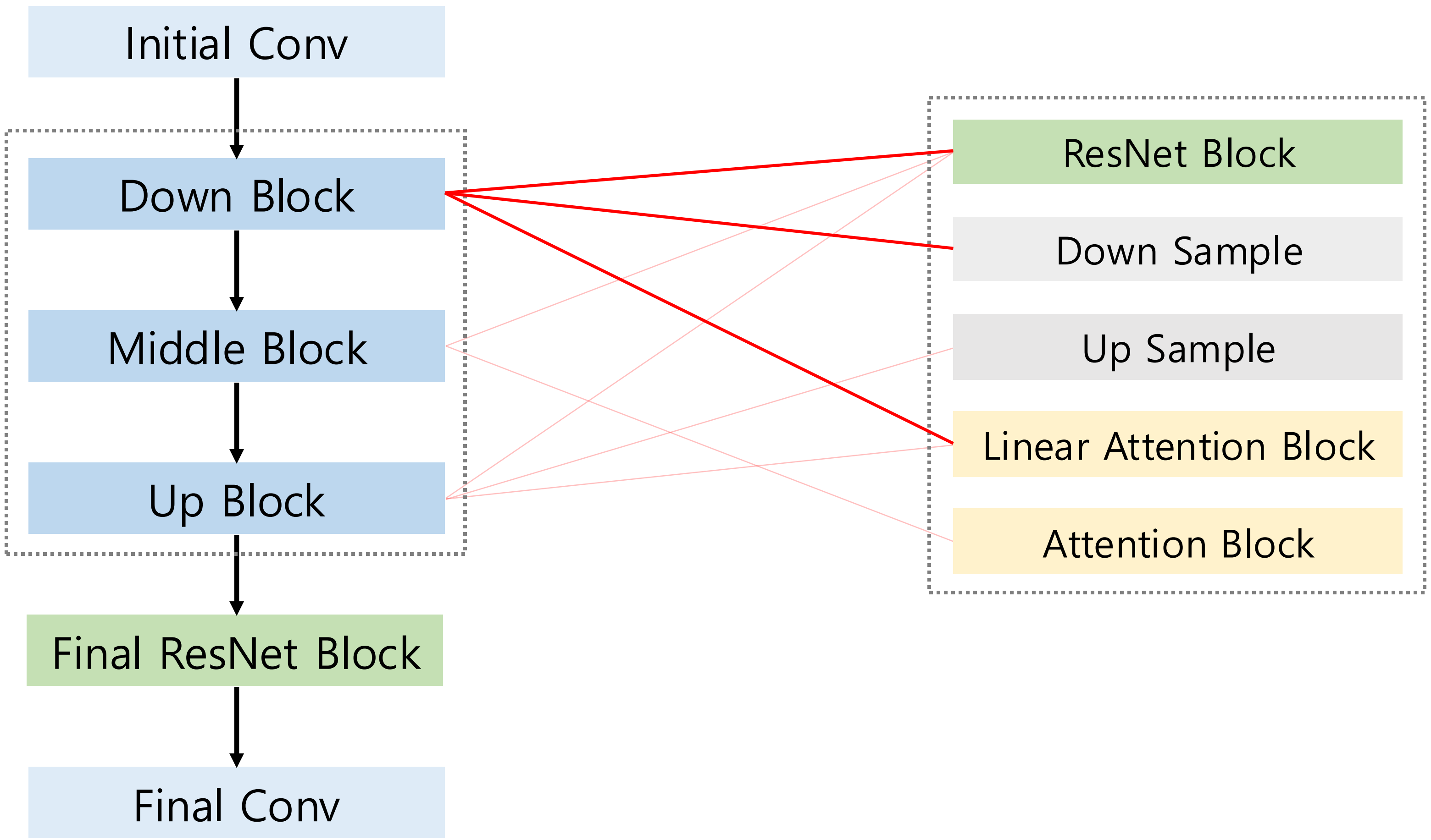
Overall Structure



Initial Conv: k=7, s=1, p=3

Final Conv: k=1, s=1, p=0

Overall Structure



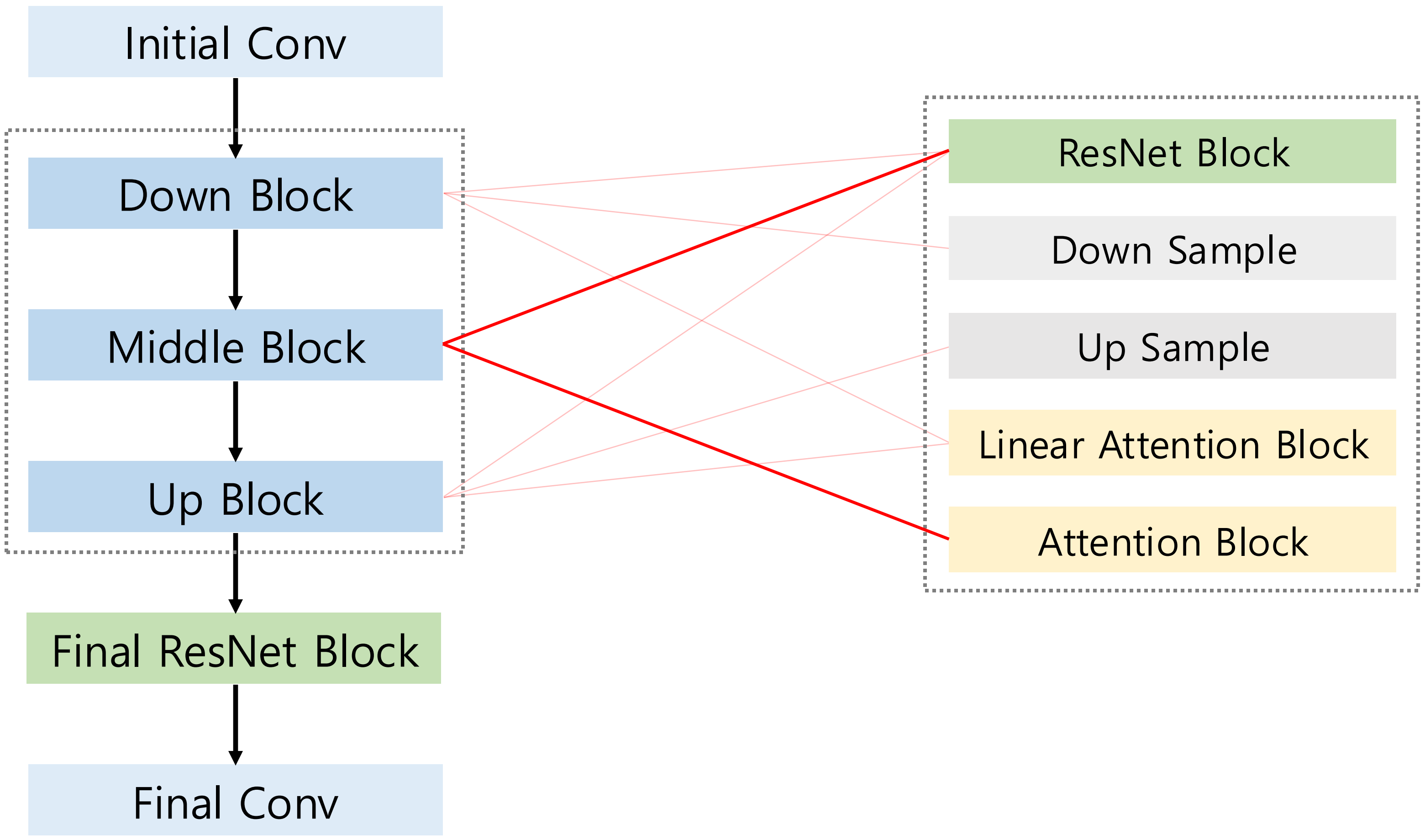
ResNet Block1

ResNet Block2

Linear Attention Block

Down Sample

Overall Structure

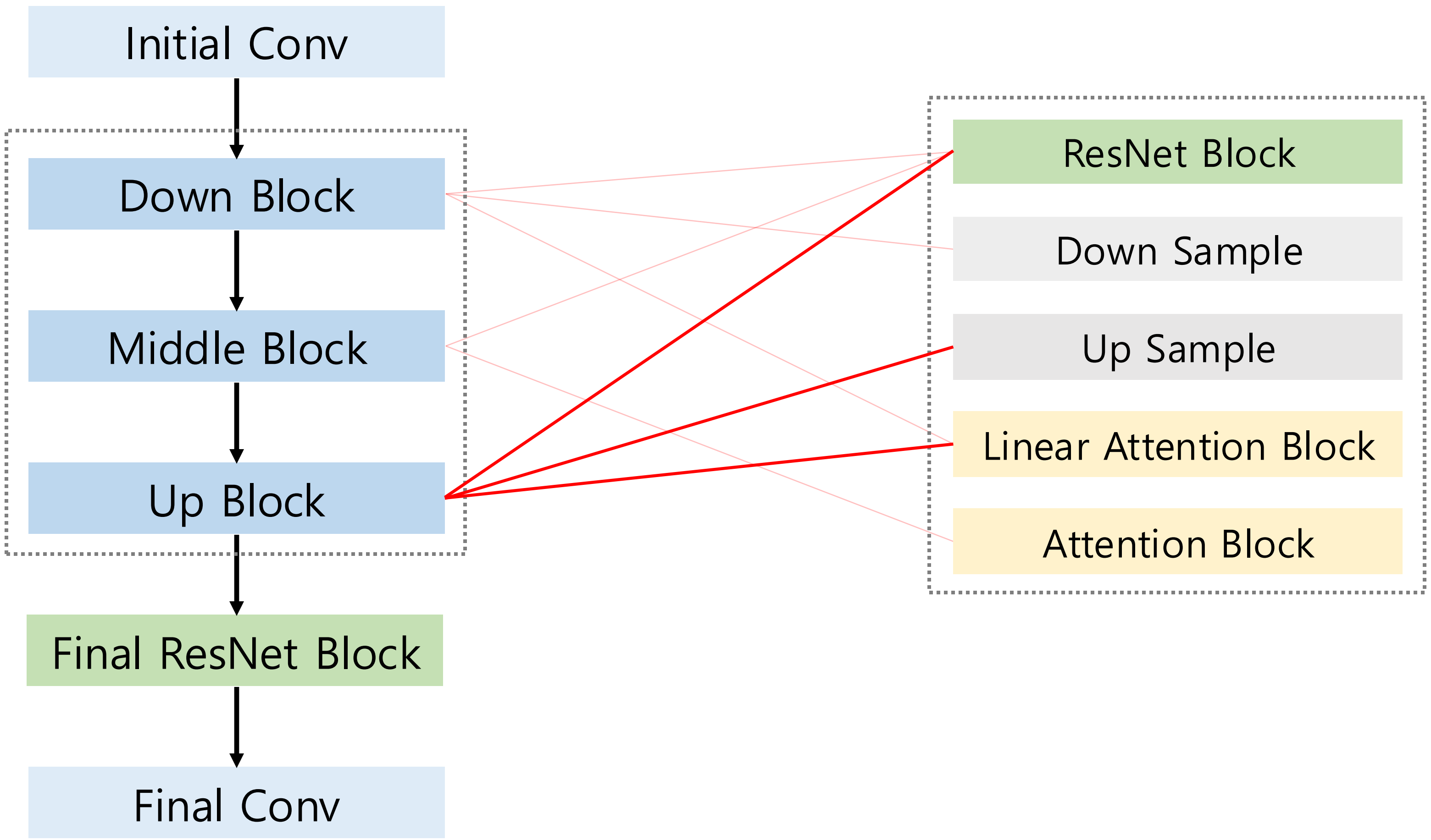


ResNet Block1

Attention Block

ResNet Block2

Overall Structure



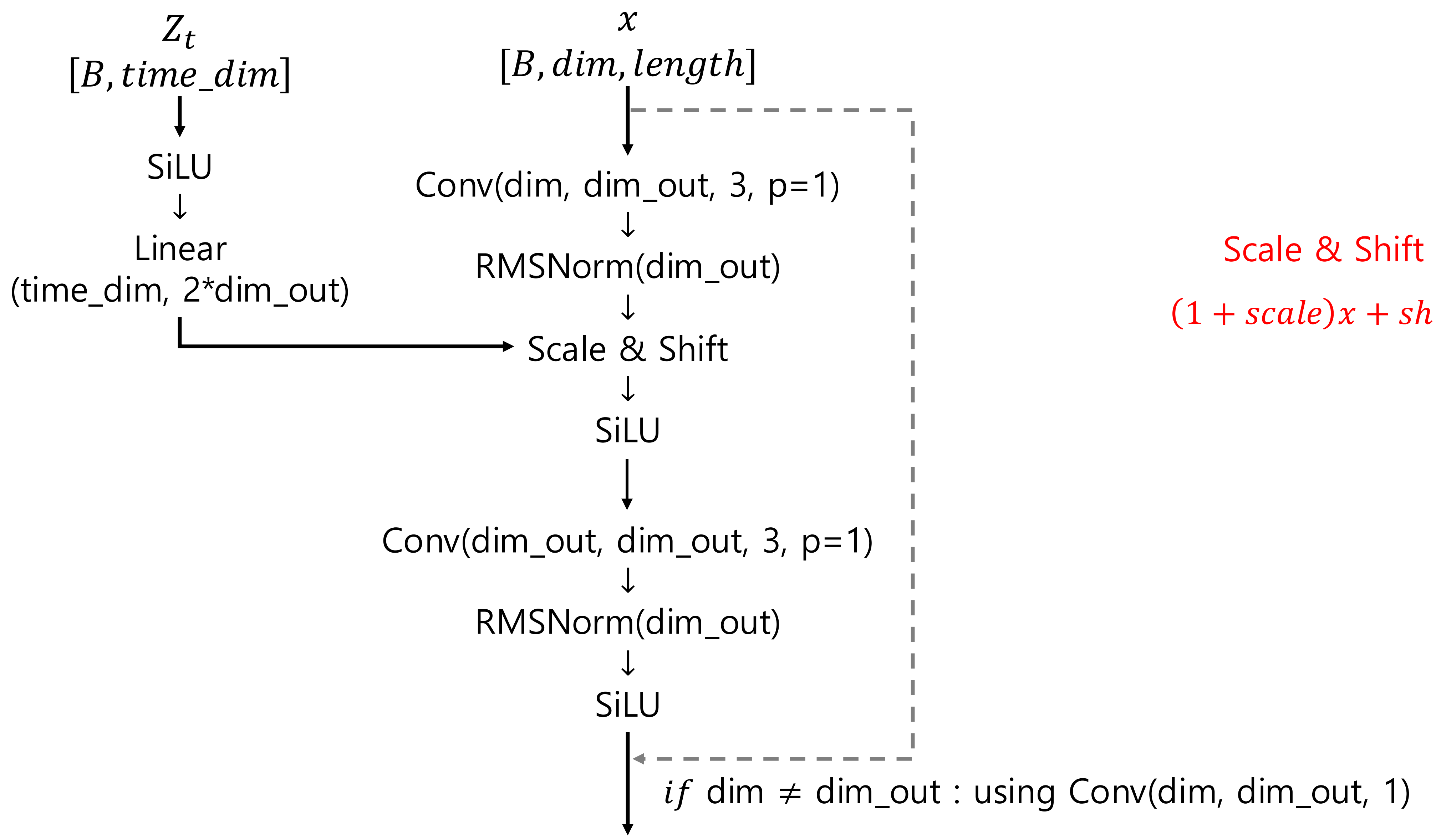
ResNet Block1

ResNet Block2

Linear Attention Block

Up Sample

ResNet Block (dim, dim_out)



Scale & Shift

$$(1 + scale)x + shift$$

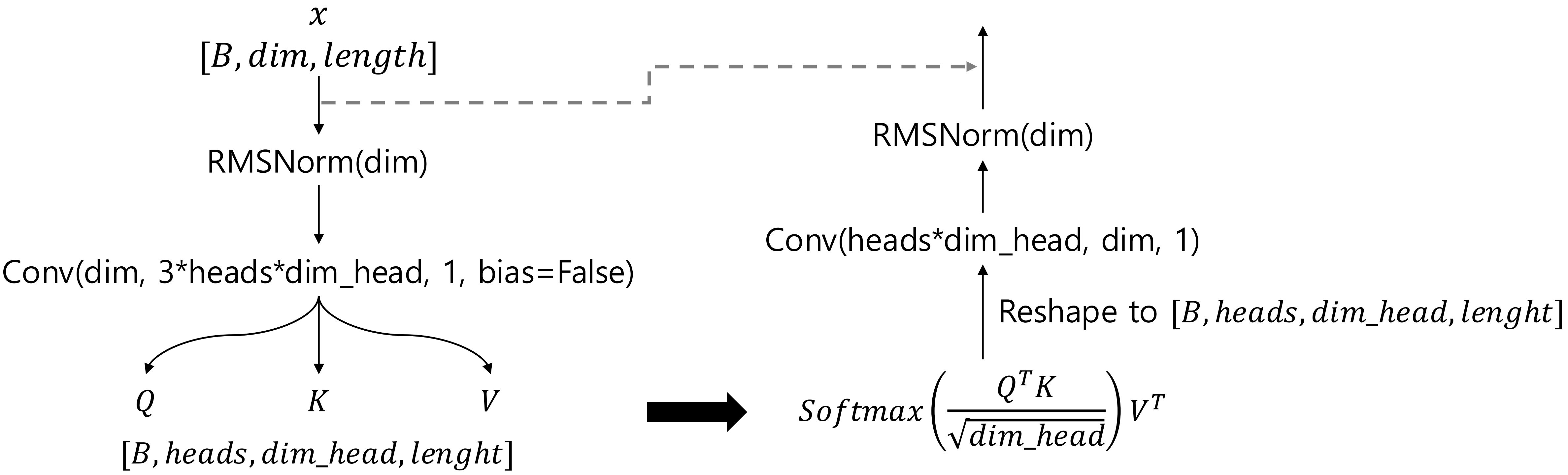
Down Sample (dim, dim_out)

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

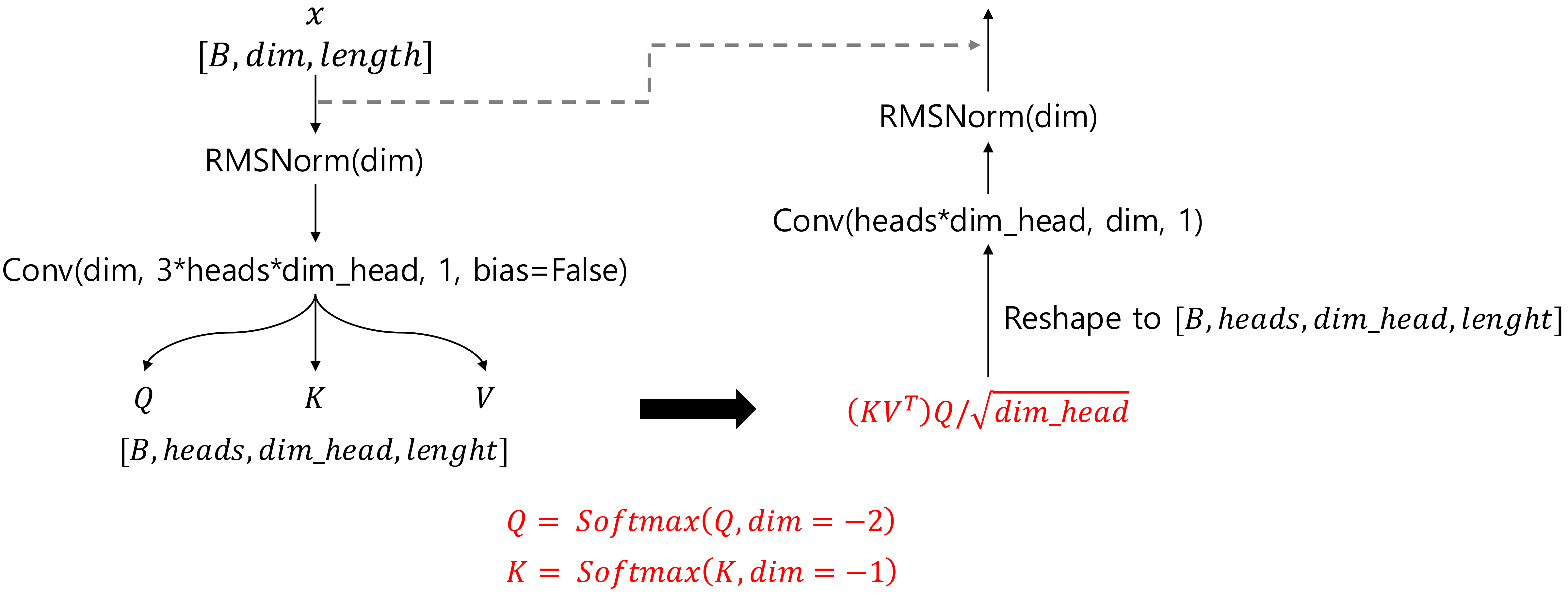
Up Sample (dim, dim_out)

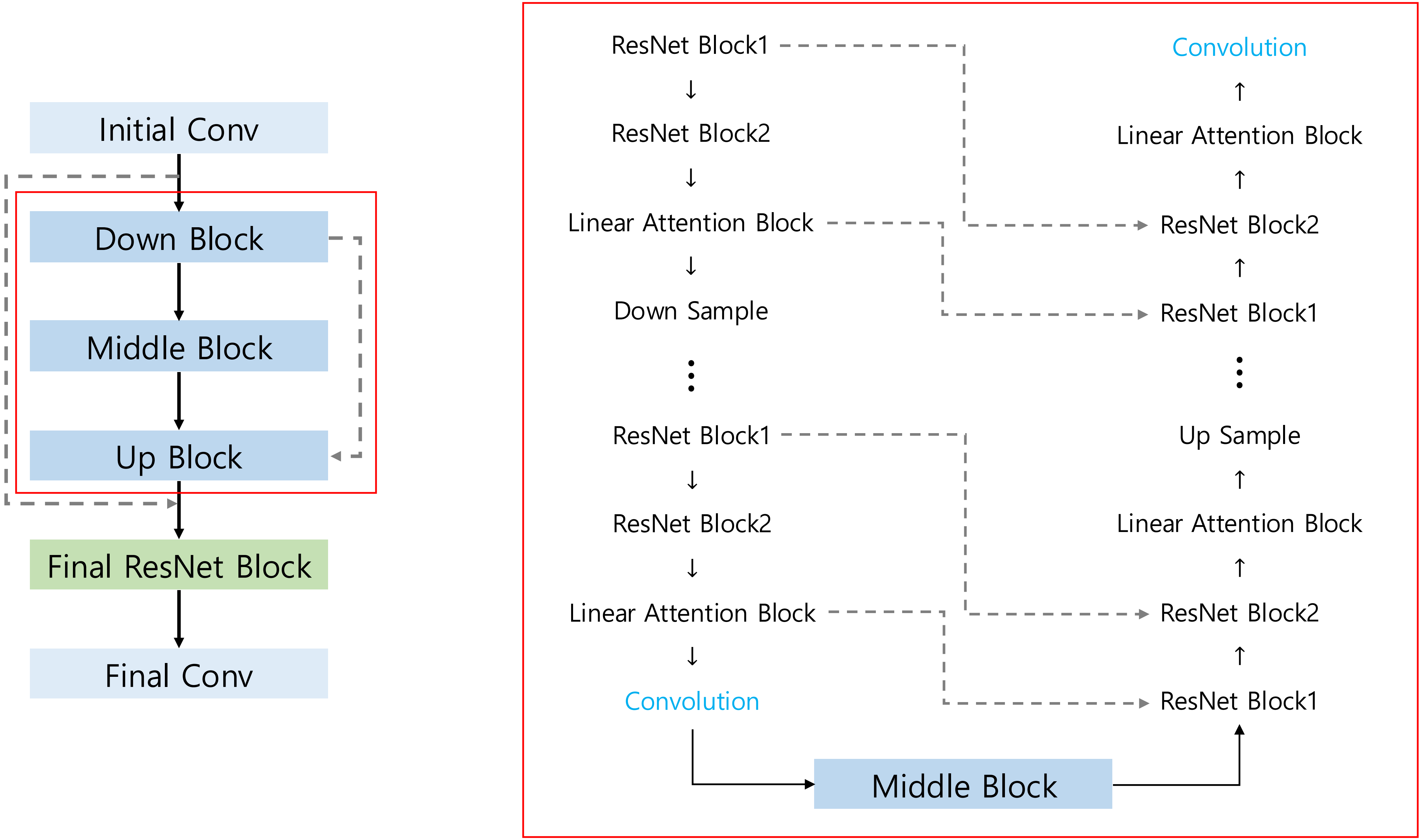
$\underset{[B, dim, length]}{x}$ $\longrightarrow$ Upsample(scale_factor=2, mode='nearest') $\longrightarrow$ Conv(dim, dim_out, 3, p=1)

Attention Block (dim, heads, dim_head)



Linear Attention Block (dim, heads, dim_head)




Initial Conv
Down Block
Middle Block
Up Block
Final ResNet Block
Final Conv
ResNet Block1
ResNet Block2
Linear Attention Block
Down Sample
ResNet Block1
ResNet Block2
Linear Attention Block
Convolution
Middle Block
Convolution
Linear Attention Block
ResNet Block2
ResNet Block1
Up Sample
Linear Attention Block
ResNet Block2
ResNet Block1

Forward Process: $X_{t-1} \rightarrow X_t \Rightarrow X_0 \rightarrow X_t$

Assumption

1. $q(X_t|X_{t-1}) \coloneqq N\left(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t \mathbf{I}\right)$
2. *Markov Chain*

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

$$\Rightarrow X_t = \sqrt{\bar{\alpha}_t}X_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon \quad where \quad \varepsilon \sim N(0, \mathbf{I})$$

Backward Process: $X_t \rightarrow X_{t-1}$

$$q(X_{t-1}|X_t, X_0) = N\left(X_{t-1}; \tilde{\mu}_t(X_t, X_0), \tilde{\beta}_t \mathbf{I}\right)$$

$$\boxed{p_\theta(X_{t-1}|X_t) = N\left(X_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(X_t, t)\right)}$$

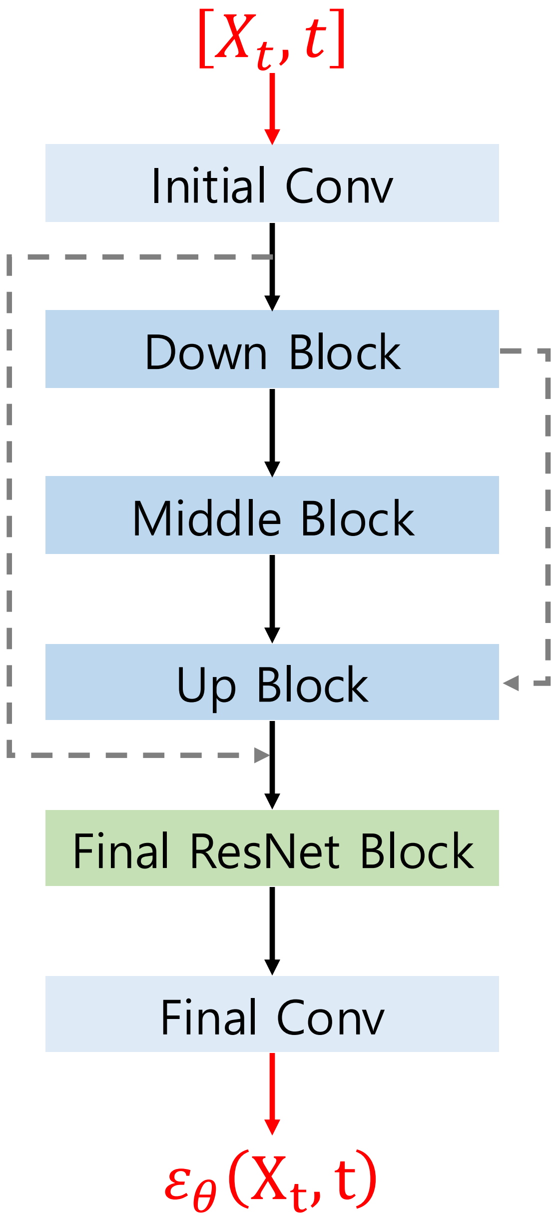
Goal of DDPM

$$q(X_{t-1}|X_t, X_0) = N\left(X_{t-1}; \tilde{\mu}_t(X_t, X_0), \tilde{\beta}_t \mathbf{I}\right)$$

$$\boxed{p_\theta(X_{t-1}|X_t) = N\left(X_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(X_t, t)\right)}$$

1. $p_\theta(X_{t-1}|X_t) \approx q(X_{t-1}|X_t, X_0) \Rightarrow D_{KL}\left(q(X_{t-1}|X_t, X_0) \parallel p_\theta(X_{t-1}|X_t)\right)$
2. *Additional Assumption:* $\Sigma_\theta(X_t, t) = \sigma_t^2 \mathrm{I}$

$$\boxed{p_\theta(X_{t-1}|X_t) \approx q(X_{t-1}|X_t, X_0) \Rightarrow \min\|\varepsilon_\theta(X_t, t) - \varepsilon\|}$$



Training Process

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

2,3,4 $\Rightarrow X_0 \rightarrow X_t$ (Forward Process)

5 $\Rightarrow \varepsilon_\theta(X_t, t) \approx \varepsilon$ (Backward Process)