Algorithm 3 EfficientDet Backbone Algorithm

Require: Input Image I, Scaling factor α, Down-sampling rate R, Depthwise convolution kernel size D, Expansion factor T, Reduction factor T'

Ensure: Feature map with fixed resolution

function EfficientDetBackbone(I, α, R, D, T, T')

$S \leftarrow$ minimum dimension of I

$resolution \leftarrow \lceil \alpha \times S \ / \ R \rceil$

$X \leftarrow$ input feature map of size $H \times W \times C$

$K \leftarrow$ number of channels in X

$X_d \leftarrow$ depthwise convolution on X with kernel size $K \times D \times D$:

$\forall i, j, k : X_d(i, j, k) = \sum_{u,v} X(i + u, j + v, k) \times W_d(u, v, k)$

where W_d is the depthwise convolution kernel of size $D \times D \times K$

$K' \leftarrow K \times T$

$X_p \leftarrow$ pointwise convolution on X_d with kernel size $1 \times 1 \times K'$:

$\forall i, j, k' : X_p(i, j, k') = \sum_k X_d(i, j, k) \times W_p(1, 1, k, k')$

where W_p is the pointwise convolution kernel of size $1 \times 1 \times K \times K'$

$X_p \leftarrow$ swish activation function on X_p with trainable parameter β:

$\forall i, j, k' : X_p(i, j, k') = \frac{X_p(i,j,k')}{1+\exp(-(\beta \times X_p(i,j,k')))}$

$X_{\text{out}} \leftarrow$ pointwise convolution on X_p with kernel size $1 \times 1 \times T'$:

$\forall i, j, k : X_{\text{out}}(i, j, k) = \sum_{k'} X_p(i, j, k') \times W_{\text{out}}(1, 1, k', k)$

where W_{out} is the pointwise convolution kernel of size $1 \times 1 \times K' \times T'$

$X_{\text{out}} \leftarrow$ elementwise addition of X and X_{out}:

$\forall i, j, k : X_{\text{out}}(i, j, k) = X(i, j, k) + X_{\text{out}}(i, j, k)$

$X_{\text{out}} \leftarrow$ ReLU activation function on X_{out}:

$\forall i, j, k : X_{\text{out}}(i, j, k) = \max(0, X_{\text{out}}(i, j, k))$

return feature map with resolution $resolution$

end function