

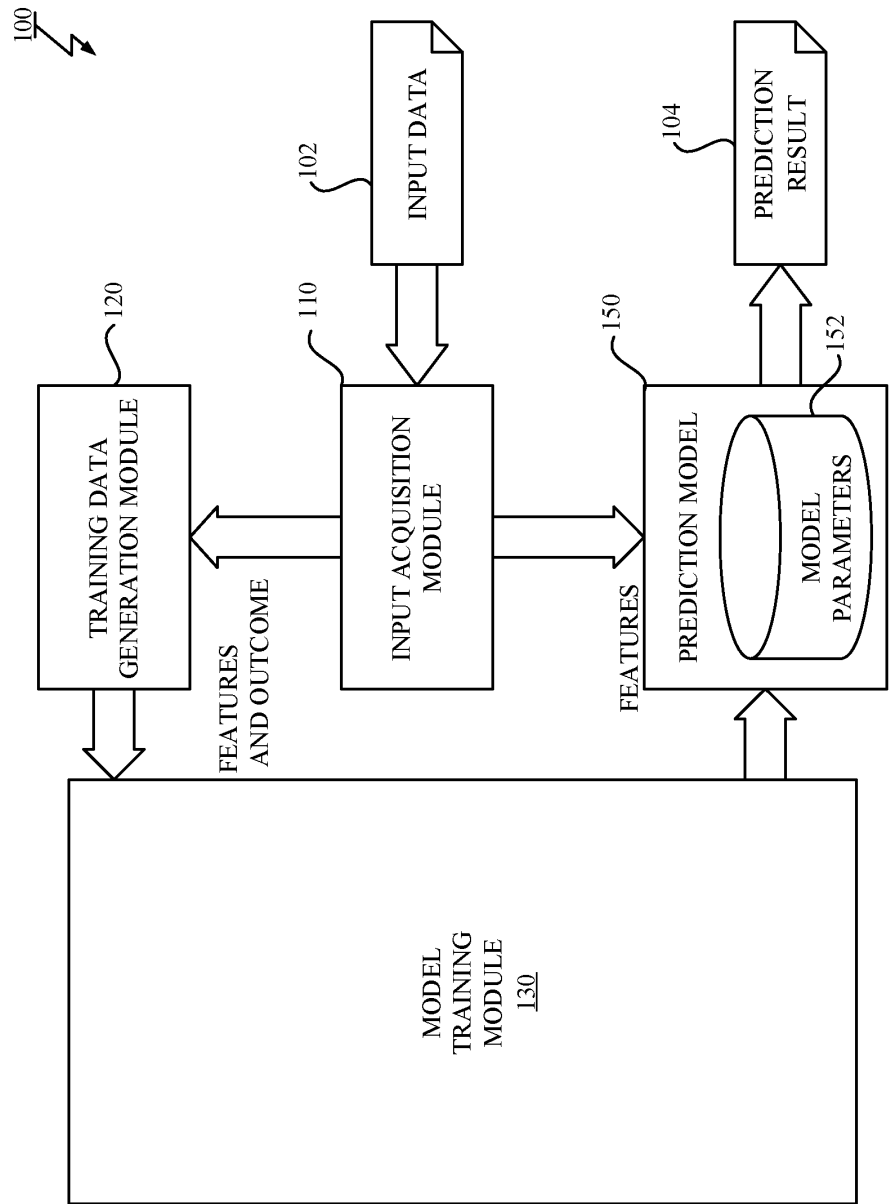


FIG. 1

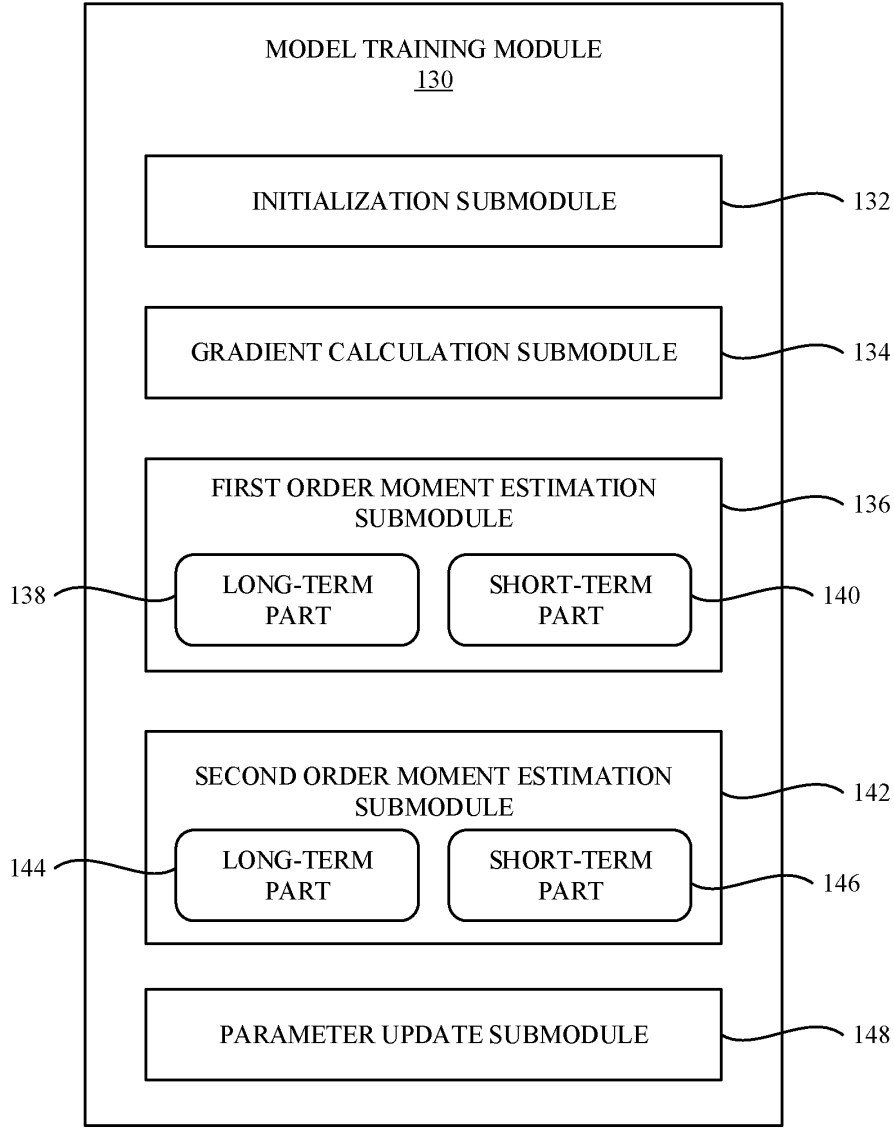


FIG. 2

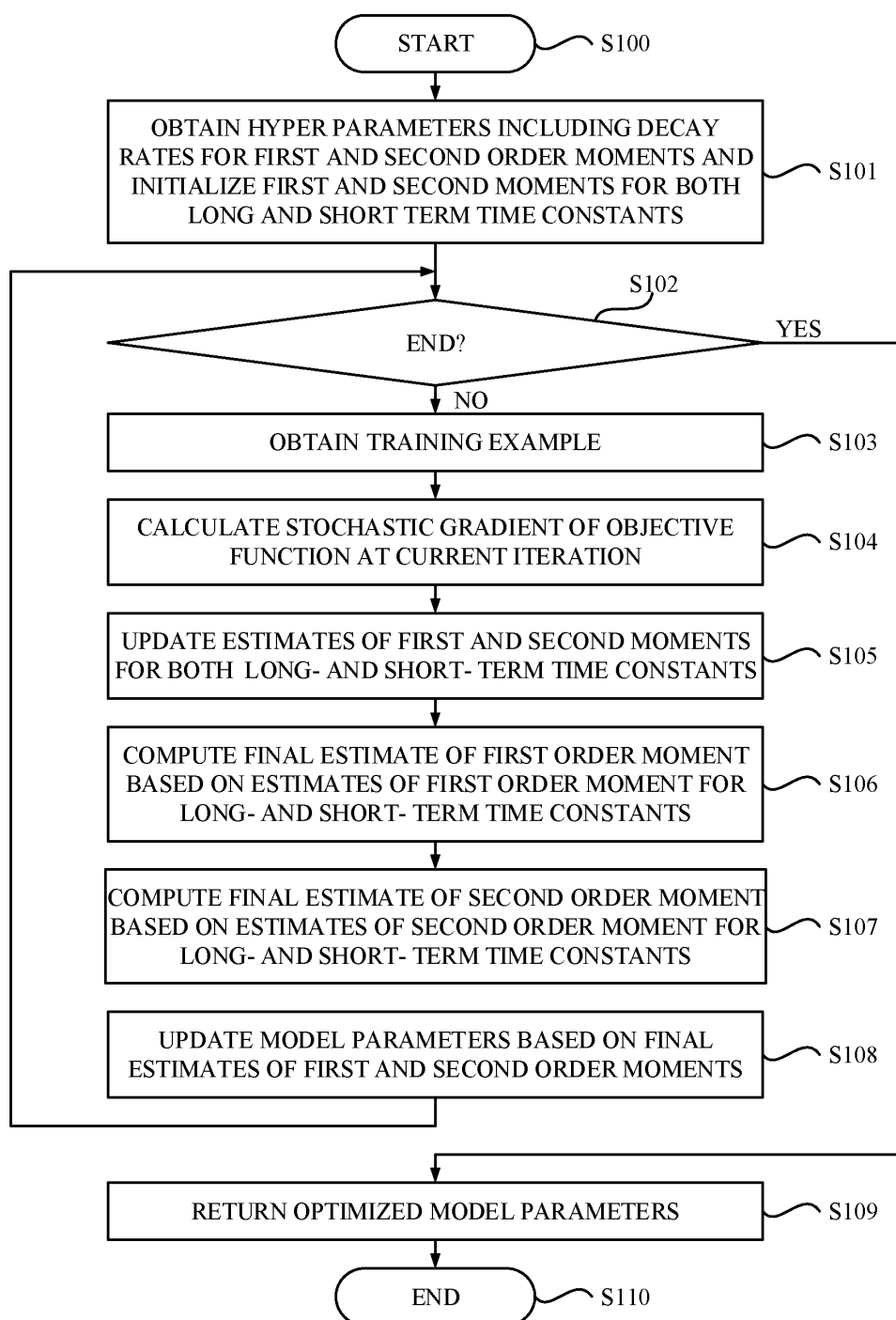


FIG. 3

```

010 Require:  $r$ : base step size or learning rate hyper-parameter
020 Require:  $bl_1, bl_2 \in [0, 1]$ : decay rates for first and second order moment estimates
030 Require:  $f(p)$ : objective function with model parameters  $p$ 
040 Require:  $p_0$ : initial model parameters vector
050 Require:  $d1, d2 > 0$ : powers of changing the decay rates  $bl_1, bl_2$ 
060  $t \leftarrow 0$  (initialize iteration step)
070  $ml_0 \leftarrow 0$  (initialize first order moment vector for long time constant)
080  $vl_0 \leftarrow 0$  (initialize second order moment vector for long time constant)
090  $bs_1 \leftarrow bl_1^{d1}$  (set decay rate of short term time constant for first order moment estimate)
100  $bs_2 \leftarrow bl_2^{d2}$  (set decay rate of short term time constant for second order moment estimate)
110  $ms_0 \leftarrow 0$  (initialize first order moment vector for short time constant)
120  $vs_0 \leftarrow 0$  (initialize second order moment vector for short time constant)
130 while  $p_t$  not converged do
140    $t \leftarrow t + 1$ 
150    $g_t \leftarrow \nabla_p f(p_{t-1})$  (get stochastic gradient of the objective function at time step  $t$ )
160    $ml_t \leftarrow bl_1 \bullet ml_{t-1} + (1 - bl_1) \bullet g_t$  (update first order moment vector for long time constant)
170    $vl_t \leftarrow bl_2 \bullet vl_{t-1} + (1 - bl_2) \bullet g_t^2$  (update second order moment vector for long time constant)
180    $ms_t \leftarrow bs_1 \bullet ms_{t-1} + (1 - bs_1) \bullet g_t$  (update first order moment vector for short time constant)
190    $vs_t \leftarrow bs_2 \bullet vs_{t-1} + (1 - bs_2) \bullet g_t^2$  (update second order moment vector for short time constant)
200    $M_t \leftarrow I(ml_t \circ ms_t \geq 0) \circ H(ml_t / (1 - bl_1'), ms_t / (1 - bs_1'))$  (compute bias corrected first order moment vector)
210    $V_t \leftarrow \max(vl_t / (1 - bl_2'), vs_t / (1 - bs_2'))$  (compute bias corrected second order moment vector)
220    $p_t \leftarrow p_{t-1} - r \bullet M_t / (\sqrt{V_t} \cdot e)$  (update model parameters)
230 end while
240 return  $p_t$ 

```

FIG. 4

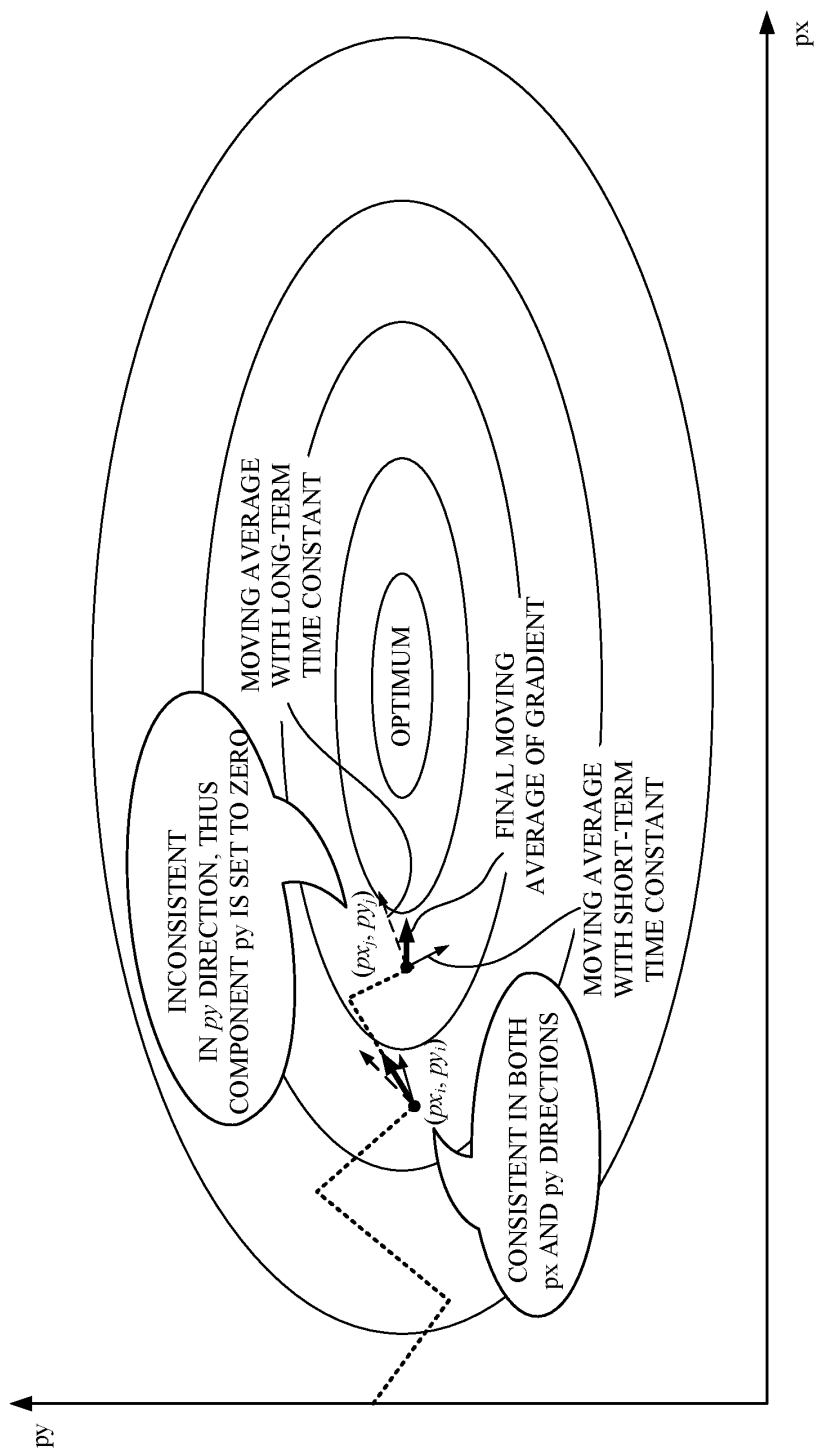


FIG. 5

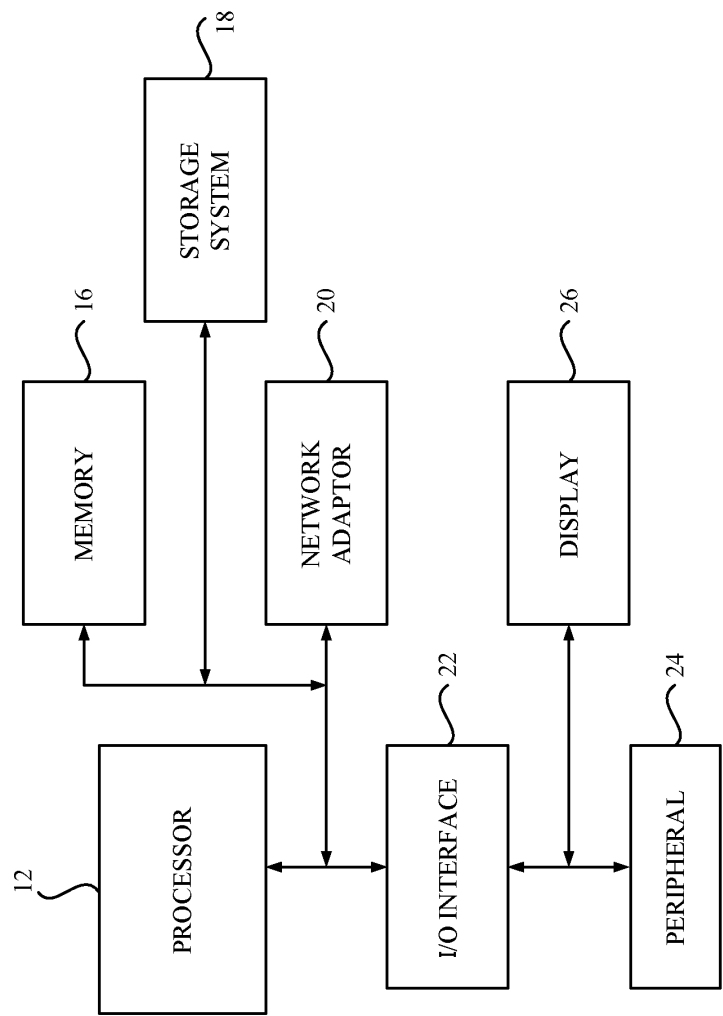


FIG. 6