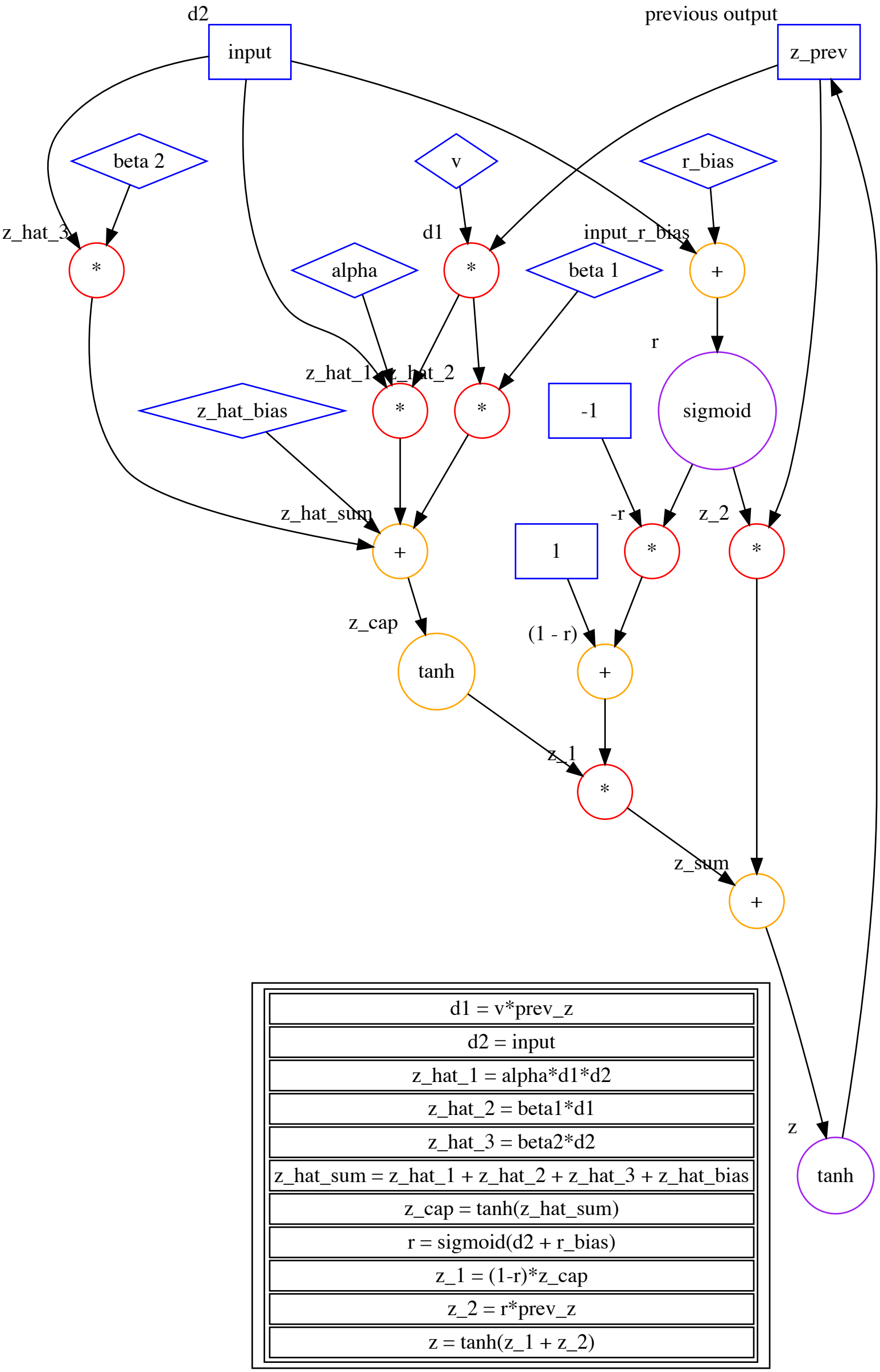
d2
input
previous output
z_prev
beta 2
v
r_bias
z_hat_3
*
d1
input_r_bias
alpha
*
beta 1
+
r
z_hat_1
z_hat_2
z_hat_bias
*
*
-1
sigmoid
z_hat_sum
-r
z_2
+
1
*
*
z_cap
(1 - r)
tanh
+
z_1
*
z_sum
+
z
tanh
d1 = v*prev_z
d2 = input
z_hat_1 = alpha*d1*d2
z_hat_2 = beta1*d1
z_hat_3 = beta2*d2
z_hat_sum = z_hat_1 + z_hat_2 + z_hat_3 + z_hat_bias
z_cap = tanh(z_hat_sum)
r = sigmoid(d2 + r_bias)
z_1 = (1-r)*z_cap
z_2 = r*prev_z
z = tanh(z_1 + z_2)