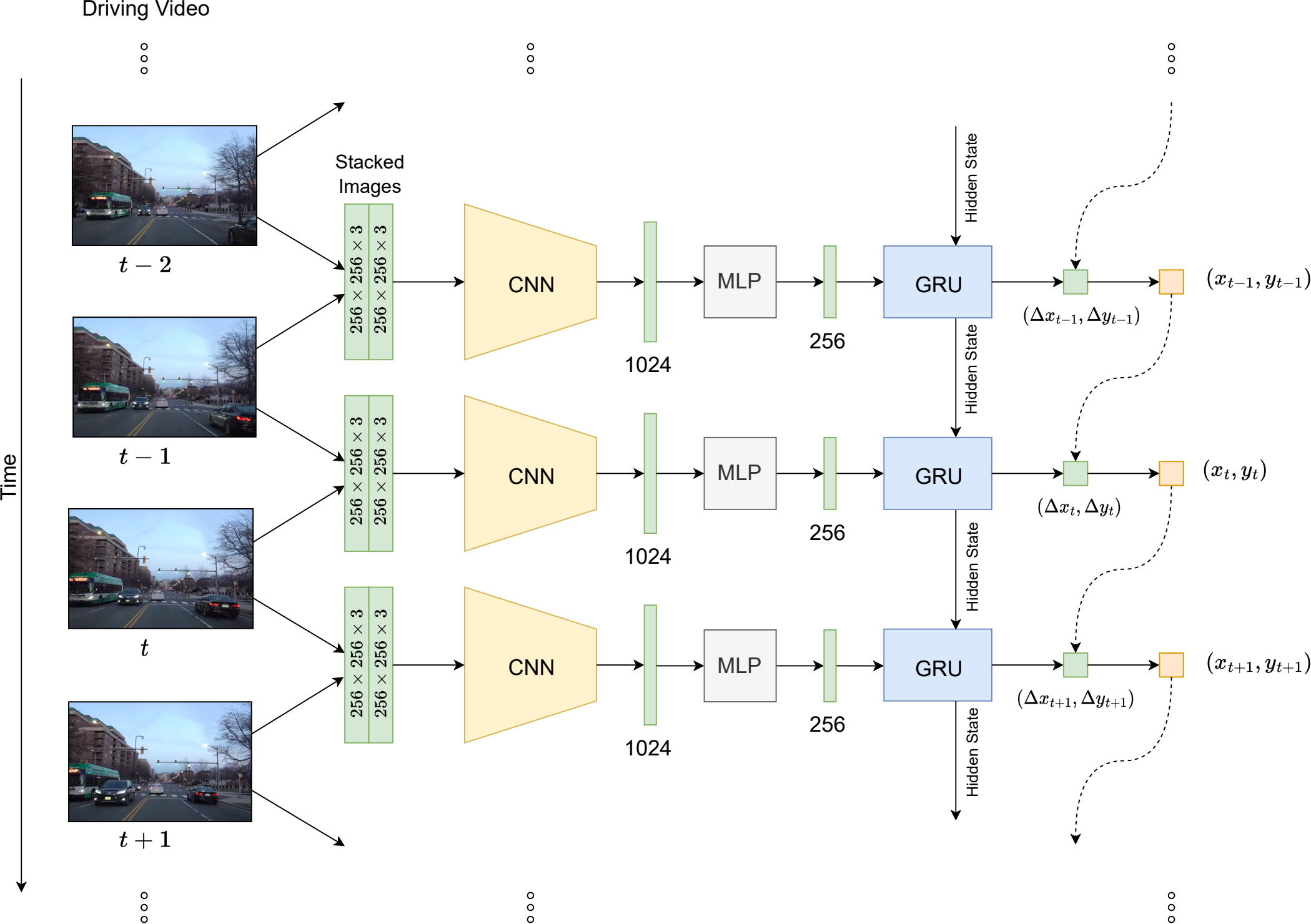
Driving Video
Time
$t-2$
$t-1$
t
$t+1$
Stacked Images
$256 \times 256 \times 3$
$256 \times 256 \times 3$
CNN
1024
MLP
256
GRU
Hidden State
$(\Delta x_{t-1}, \Delta y_{t-1})$
$(\Delta x_t, \Delta y_t)$
$(\Delta x_{t+1}, \Delta y_{t+1})$
(x_{t-1}, y_{t-1})
(x_t, y_t)
(x_{t+1}, y_{t+1})