GAE in case of trajectories with actions of variable length

In general, timesteps in general advantage estimation (GAE) [High Dimensional Continuous Control Using Generalized Advantage Estimation, Schulman et al. 2016 and Proximal Policy Optimization Algorithms, Schulman et al. 2017] are assumed to be equidistant. This might mean that actions executed in an RL environment all take the same time and a learning agent equally steps through the environment.
In cases where this assumption doesn't hold (e.g. actions might take from 1 second up to some minutes), here is a proposal to adapt GAE to cases where actions can have a variable length and agent's steps are not equidistant (perhaps this holds in cases which are also known as Semi-MDPs).

Preliminaries:
Let's denote the time between two agent's actions (i.e. primitive step from t_i to t_{i+1}) by

$$\Delta_{t_i} = t_{i+1} - t_i$$

Furthermore, let

$$R_{t_i} = \sum_{l=0}^{\Delta_{t_i}-1} \gamma^l \cdot r_{t_i+l}$$

be the "accumulated discounted reward/return" collected between primitive steps t_i and t_{i+1}, where each r_{t_i+l} is a single reward collected by an agent on an "inner, finer equidistant time grid" (for simplification we might assume $l = 0, 1, 2, \dots, \Delta_{t_i} - 1$). Many of these r_{t_i+l} can be zero, in the simplest case just r_{t_i} is different from zero.

Adapted GAE formula:
A proposal for an adapted GAE formula (n-step trajectories) might look like this:

$$\hat{A}_{t_i} = \delta_{t_i} + (\gamma\lambda)^{t_{i+1}-t_i} \cdot \delta_{t_{i+1}} + (\gamma\lambda)^{t_{i+2}-t_i} \cdot \delta_{t_{i+2}} + \cdots + (\gamma\lambda)^{t_{n-1}-t_i} \cdot \delta_{t_{n-1}}$$

with the TD residual

$$\delta_{t_i} = \sum_{l=0}^{\Delta_{t_i}-1} \gamma^l \cdot r_{t_i+l} + \gamma^{\Delta_{t_i}} \cdot V\left(s_{t_i+\Delta_{t_i}}\right) - V\left(s_{t_i}\right) = R_{t_i} + \gamma^{t_{i+1}-t_i} \cdot V\left(s_{t_{i+1}}\right) - V\left(s_{t_i}\right)$$

for $i = 0, 1, \dots, n-1$.

"accumulated discounted reward/return" should be provided from environment

E.g.,

$$\begin{aligned}\hat{A}_{t_0} &= \delta_{t_0} + (\gamma\lambda)^{\Delta_{t_0}} \cdot \delta_{t_0+\Delta_{t_0}} + (\gamma\lambda)^{t_2-t_0} \cdot \delta_{t_2} + \cdots + (\gamma\lambda)^{t_{n-1}-t_0} \cdot \delta_{t_{n-1}} \\ &= \delta_{t_0} + (\gamma\lambda)^{t_1-t_0} \cdot \delta_{t_1} + (\gamma\lambda)^{t_2-t_0} \cdot \delta_{t_2} + \cdots + (\gamma\lambda)^{t_{n-1}-t_0} \cdot \delta_{t_{n-1}}\end{aligned}$$

In contrary to standard GAE, this adapted GAE needs the "accumulated discounted rewards/returns" R_{t_i} (instead of r_t) and additionally the actual timestamps t_i.