# A Soft Actor-Critic algorithm as implemented for TractOracle-RL

**Algorithm 1** TractOracle-RL implementation of Soft Actor-Critic [12]

```
Require: $\theta, \phi_1, \phi_2$          ▷ Parameters for the actor and twin critics, respectively
Require: $\alpha$                          ▷ Entropy parameter for the actor
Require: $\phi'_1, \phi'_2$                ▷ Target parameters
Require: $D$                               ▷ Replay buffer, initialized empty
Require: $E_{s_0}$                         ▷ Environment in initial state
Require: $\mathcal{H}$                     ▷ Desired minimum entropy of $\pi$
Require: $\eta$                            ▷ Learning rate
  for $i = 0..1000$ do
    $t \leftarrow 0$
    $s_t \leftarrow E_{s_0}$               ▷ Initial state from the environment
    while $t \neq T$ do
      $a_t \sim \pi_\theta(s_t)$
      $s_{t+1}, r_t, \mathbb{1}_t \sim E_{s_t}(a_t)$          ▷ Apply action to state of environment
      $D \leftarrow D \cup (s_t, a_t, r_t, s_{t+1}, \mathbb{1}_t)$          ▷ Store experience in replay buffer
      $\{(s_k, a_k, r_k, s_{k+1}, \mathbb{1}_k)\} \leftarrow D$  ▷ Sample a batch of experiences from the buffer
      $\alpha \leftarrow \alpha - \eta \nabla_\alpha \mathbb{E}_{a \sim \pi_\theta(s_k)}[\alpha \log \pi(a|s_k) - \alpha \mathcal{H}]$      ▷ Update entropy parameter
      $\theta \leftarrow \theta - \eta \nabla_\theta \mathbb{E}_{s_k}[\alpha \log \pi(a_k|s_k) - \min(Q_{\phi_1}(s_k, a_k), Q_{\phi_2}(s_k, a_k))]$
      $a_{k+1} \leftarrow \pi_\theta(s_{k+1})$
      $\hat{Q} \leftarrow r_k + \gamma \mathbb{1}_k \min(Q_{\phi'_1}(s_{k+1}, a_{k+1}), Q_{\phi'_2}(s_{k+1}, a_{k+1})) - \alpha \log \pi_\theta(a_{k+1}|s_{k+1})$
      $\phi_1 \leftarrow \phi_1 - \eta \nabla_{\phi_1} \mathbb{E}_{s_k, a_k \sim D}[\frac{1}{2}(Q_{\phi_1}(s_k, a_k) - \hat{Q})^2]$
      $\phi_2 \leftarrow \phi_2 - \eta \nabla_{\phi_1} \mathbb{E}_{s_k, a_k \sim D}[\frac{1}{2}(Q_{\phi_2}(s_k, a_k) - \hat{Q})^2]$
      $\phi'_1 \leftarrow \tau \phi_1 + (1 - \tau)\phi'_1$
      $\phi'_2 \leftarrow \tau \phi_2 + (1 - \tau)\phi'_2$
      if $\mathbb{1}$ is 1 then                ▷ If trajectory is over
        $t \leftarrow T$
      else
        $s_t \leftarrow s_{t+1}$
      end if
    end while
  end for
```

## B Additional bundles of subject 1006 of the Tractoinferno dataset

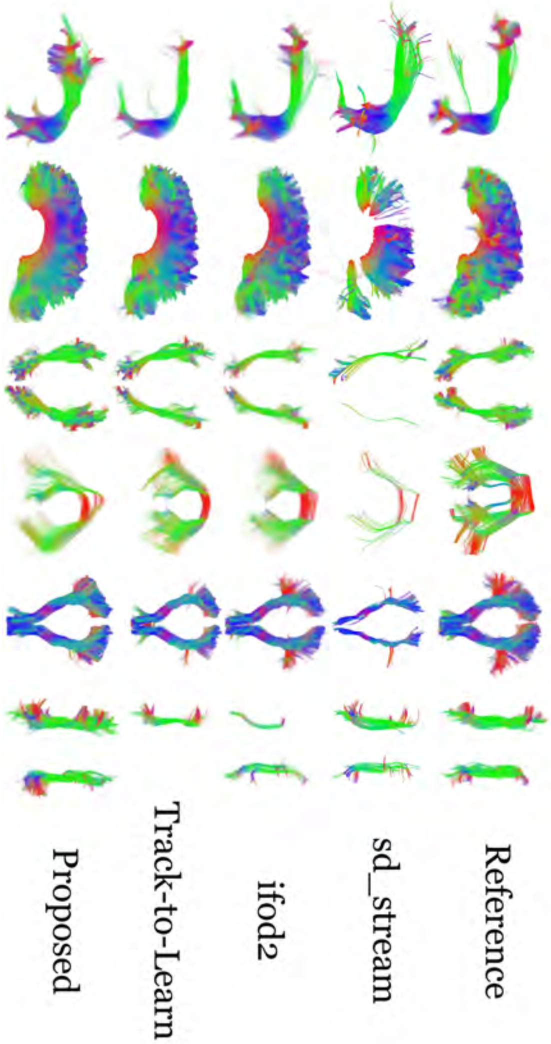

Fig. 3: Visualization of the left arcuate fasciculus (1st row), the corpus callosum (2nd row), the left-right inferior longitudinal fasciculus (3rd row), the middle cerebellar peduncle (4th row), the pyramidal tracts (fifth row) and the superior longitudinal fasciculus (sixth row) from subject 1006 of the Tractoinferno dataset (reference) and as reconstructed by all methods considered in this work.