Sarah Bechtle\*, Artem Molchanov\*, Yevgen Chebotar\*, Edward Grefenstette, Ludovic Righetti, Gaurav Sukhatme and Franziska Meier



Intelligent Systems USCUniversity of Southern California research VICUNIVERSITY OF

## $heta_{new} = h_\psi( heta, abla_ heta \mathcal{L}_\phi(g, f_ heta(x)))$

 $egin{aligned} & heta_{new} = h_\psi( heta, 
abla_ heta \mathcal{L}_\phi(g, f_ heta(x))) \end{aligned}$ 

 $( heta_{new}) = h_\psi( heta, 
abla_ heta \mathcal{L}_\phi(g, f_ heta(x)))$ 

 $( heta_{new}) = h_\psi( heta, 
abla_ heta \mathcal{L}_\phi(g, f_ heta(x)))$ 

 $( heta_{new}) = (h_{\psi}( heta, 
abla_{ heta} \mathcal{L}_{\phi}(g, f_{ heta}(x)))$ 

Majority of **Deep learning**  $= (h_{\psi}( heta, 
abla_{ heta}(\mathcal{L}_{\phi}(g, f_{ heta}(x))))$  $\theta_{new}$ 















**Learning Loss Functions** 



- Learning Loss Functions
- No loss engineering



#### **Learning Loss Functions**

- No loss engineering
- Include additional information















Ground truth label (regression/classification)



Ground truth label (regression/classification)

Final goal position (MBRL)



- Ground truth label (regression/classification)
- Final goal position (MBRL)
  - Sampled reward (MFRL)

















## x -Fully differentiable $\theta_{new} = \theta_{old} - \alpha \nabla_{\theta} \mathcal{L}_{\phi}$ , loss learning loop $\phi \leftarrow \phi - \eta abla_{\phi} \mathcal{L}_{task}(g, f_{ heta_{norm}})$ LdA

## Using the Learned Loss at Test Time

#### Using the Learned Loss at Test Time



























































#### **Shaping Loss During Training**

 $\mathcal{L}_{task}(g, f_{ heta_{new}}, *)$ 

#### **Shaping Loss During Training**

 $\mathcal{L}_{task}(g, f_{ heta_{new}}, st)$ 

 $f_{\theta} = sin(\theta x)$ 

 $f_{\theta} = sin(\theta x)$ 

 $\mathcal{L}_{\text{task}} = (\theta - \theta^*)^2$ 







#### Shaped ML3 Loss Landscape

$$\mathcal{L}_{ ext{task}} = ( heta - heta^*)^2$$



## **Shaping Loss During Training - Exploration Signal**

## **Shaping Loss During Training - Exploration Signal**



## **Shaping Loss During Training - Exploration Signal**



## **Reinforcement Learning with ML3**





$$\mathcal{L}_{ ext{task}} = \mathbb{E}_{\pi_{ heta_{ ext{new}}}, P}[R_{\pi_{ ext{new}}}]$$

## **Training - 2 Targets during meta train time**



#### **Testing - No Model is needed during meta test!**





AntGoal

ReacherGoal





ML<sup>3</sup> 5x ... 8x more sample efficient





**ML<sup>3</sup>** generalizes to new policy architectures

**Code** is available on our website: <u>https://sites.google.com/view/mlthree/</u> **Code** is available on our website: <u>https://sites.google.com/view/mlthree/</u>

More experiments and analysis in the paper!

**Code** is available on our website: <u>https://sites.google.com/view/mlthree/</u>

More experiments and analysis in the paper!

Thank you!