PAUMER: Patch Pausing Transformer for Semantic Segmentation

Evann Courdier *,a,b, Prabhu Teja S *,a,b, François Fleuret a,c

*Equal Contribution, aIDIAP Research Institute, bEPFL, cUniversity of Geneva, Switzerland,
Contact: evann.courdier@idiap.ch, prabhu.teja@idiap.ch

One training to pause them all!: Training to handle various run-time requirements

- For each batch, sample a layer \( l \sim U(\tau_l, L^2) \) and a patch pausing proportion \( \eta \sim U(\eta_l^1, \eta_l^n) \).
- We employ an auxiliary decoder after the operations of layer \( l \), and pause \( \tau_l \) patches with the lowest entropy.
- Training both main and aux decoder with cross entropy.

Pausing patches at multiple layers

Posterior entropy correlates with correctness

Figure: How long do we have to process each part of the image for correct predictions with a ViT-Ti backbone?

The network automatically pauses easy parts of the image while allocating more computation to the parts that correspond to boundaries, and to smaller and rarer classes in (c.3).

Research supported by the “Swiss Center for Drones and Robotics - SCoR” of the Department of Defence, Civil Protection and Sport via annuaire 5 1 under project n.920-38.

Acknowledgements