Bootstrapping Human Optical Flow and Pose
Aritro Roy Arko, James J. Little, Kwang Moo Yi
University of British Columbia

Problem:
- Generic optical flow methods (such as RAFT) perform better on humans when fine-tuned on human-centric scenes. In addition, they fail in cases of fast motion.
- Overlooked assumption in recent pose estimations works (such as METRO) is temporal consistency. Some methods take them into consideration but most leave it for the Neural Network to implicitly figure out and embed into the framework while training.

Solution:
- Make use of the tools that already exist—human pose estimators and optical flow networks—and enhance their performance by marrying the two.
- Create an iterative flow-pose-flow optimization framework for inference.
  - Idea originates from the fact that the movement of the joints, when projected in 2D, should follow the optical flow estimates at these locations.

Overall framework:

Flow module:

Flow from RAFT
Pose-based sketch
Target flow
Fine-tuned flow

Flow enhanced pose
Pose enhanced flow

Flow module:

\[ L_{\text{flow}}(\Phi_{\text{RAFT}}) = E[p(f^{T} - f^{t})] \]

- Generate rough optical flow map of the bones with help of pose estimator.
- This is overlaid on top of the estimated flow map (e.g., by RAFT).
- Target flow map \( f^{t} \) produced.
- Minimize smooth \( \ell_{1} \) norm \( \rho \) between predicted flow \( f^{t} \) and \( f^{t} \). Update parameters of the RAFT model \( \Phi_{\text{RAFT}} \) to get fine-tuned optical flow.

Pose module:

\[ L_{\text{pose}}(X, C) = L_{\text{opt}}(X, C) + L_{2D}(X, C) + L_{3D}(X, C) \]

We directly optimize the 3D joint estimates based on optical flow consistency \( L_{\text{opt}}(X, C) \), 3D joint consistency \( L_{2D}(X, C) \), 2D joint consistency \( L_{2D}(X, C) \) and temporal consistency \( L_{\text{temp}}(X, C) \).

Results:

Efficiency / Quality Trade-off
(1 Head, 4 Heads, 8 Heads, 16 Heads)

Project Page:
ubc-vision.github.io/derf

Acknowledgements:
This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, NSERC Collaborative Research and Development Grant, Google, Compute Canada, and Advanced Research Computing at the University of British Columbia.

METRO
RAFT

Ablation Study:

- Effects of adding different loss terms to our pose refinement pipeline.

(a) \( \text{MPPE vs cycles} \) (b) \( \text{EPE vs cycles} \)
(c) \( \text{MPPE vs cycles in GT} \) (d) \( \text{EPE vs cycles in GT} \)

- Pose and flow errors with respect to the number of optimization cycles.