

Video you only look once: Overall temporal convolutions for action recognition

Author: Longlong Jing, Xiaodong Yang, Yingli Tian

この論文の概要

- Video you only look once という手法の開発
- 従来の動画認識より計算量が圧倒的に少ない
- 動画の内容について、判断する事ができる



Video-YOLOの具体的な手法

1. ビデオ全体から、ランダムにフレームを選択
2. プロキシビデオの作成
3. 3D-CNNによる学習
4. 学習モデルからスコアを算出

本研究では、これを複数で比較した

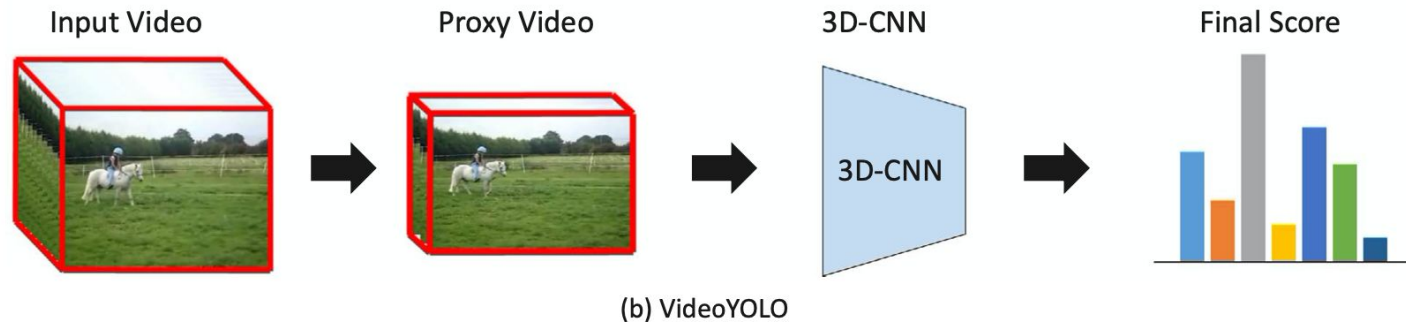
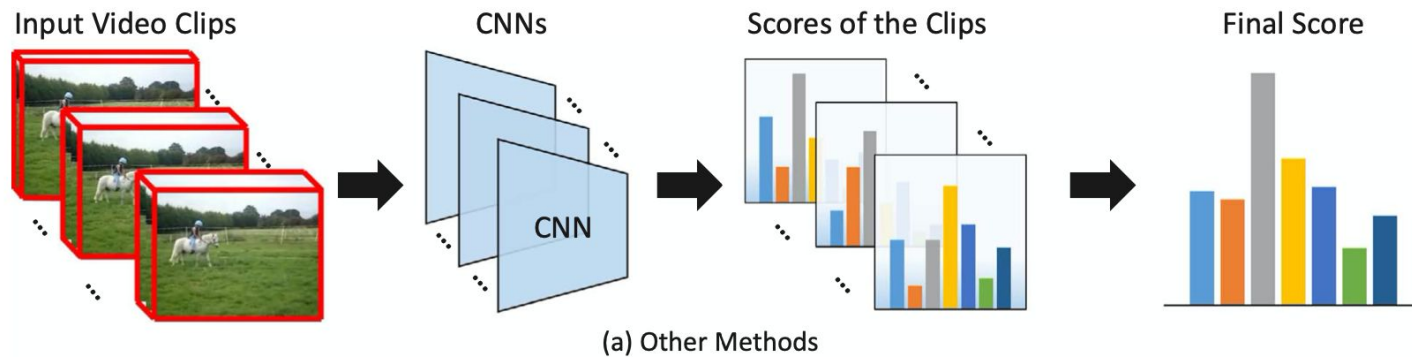


従来の手法との比較

- 従来の手法では、動画内の全てのフレームを用いて学習していた
- 動画からプロキシビデオを作成し、学習する事によって、大幅な計算コスト削減



従来の手法との比較(論文中の図)



この手法の信頼性

- プロキシ動画を作成する際の、抽出するフレーム数
- 動画内で、どのくらいの時間同じ行動を取り続けているのか

この2点がモデルの精度に大きく関わる



結論

- Video-YOLOは長時間の動画を処理する際に、非常に有効である
- Video-YOLOはポストプロセスが一切不要であり、リアルタイムな処理に適している



参考文献

Longlong Jing, Xiaodong Yang, Yingli Tian,
Video you only look once: Overall temporal
convolutions for action recognition,
Journal of Visual Communication and Image
Representation, Volume 52, 2018, Pages 58-65,
ISSN 1047-3203

