AutoVideo: An Automated Video Action Recognition System

Daochen Zha1* Zaid Pervaiz Bhat2* Yi-Wei Chen2* Yicheng Wang2* Sirui Ding2* Jiaben Chen3* Kwei-Herng Lai1*
Mohammad Qazim Bhat2* Anmoll Kumar Jain2 Alfredo Costilla Reyes1 Na Zou4 Xia Hu1

Department of Computer Science, Rice University

Introduction

Given a collection of video clips with labelled human actions, AutoVideo provides a Python interface for AutoML in video action recognition with an exhaustive list of 188 primitives alongside a Graphical User Interface (GUI).

Motivation

While prior research has achieved promising results for action recognition, they heavily rely on human efforts:

- Complexity: We often need a very complex training pipeline, including but not limited to data loading, frame extraction, video cropping/scaling, video augmentation, model training, etc.
- Laborious Tuning: We often need extensive laborious trials on different combinations of the modules and their hyperparameters.

We are motivated to simplify and automate this process.

Challenges

- How can we design a unified and easy-to-use interfaces for all the primitives for action recognition?
- How can we allow users to easily construct a pipeline and perform AutoML search?

AutoVideo System

To address the challenges, we propose AutoVideo: An Automated Video Action Recognition System, a highly modular and easy-to-use framework for automated video action recognition. AutoVideo is featured for:

- **Modularity.** We provide a highly modular and extendable infrastructure following the standard pipeline language.

- **Pipeline Construction** We support an exhaustive list of 188 primitives for pipeline construction.

- **Pipeline Search.** Use data-driven tuners to save the efforts of pipeline tuning.

- **GUI.** We provide an easy-to-use Graphical User Interface (GUI) enabling users to manually construct pipelines in a drag-and-drop fashion.