Nuisance-Label Supervision: Robustness Improvement by Free Labels

Xinyue Wei, Weichao Qiu, Yi Zhang, Zihao Xiao, Alan Yuille
University of California San Diego, Johns Hopkins University

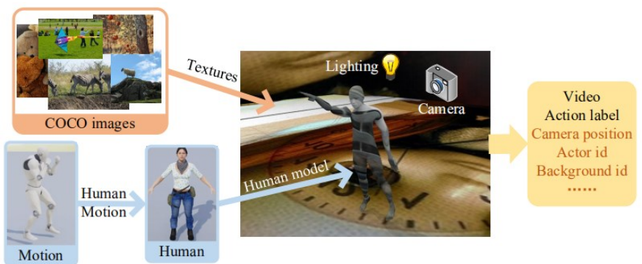
2021 ICCV OCTOBER 11-17 VIRTUAL

Introduction

Many factors can affect model robustness, and spurious correlation is one of them. Spurious correlation refers to using inaccurate information to make predictions.
We present Nuisance-label Supervision (NLS) module to make models more robust to irrelevant information, i.e. nuisance factor variations. We propose three practical and low-cost ways to collect the labels of nuisance factors.

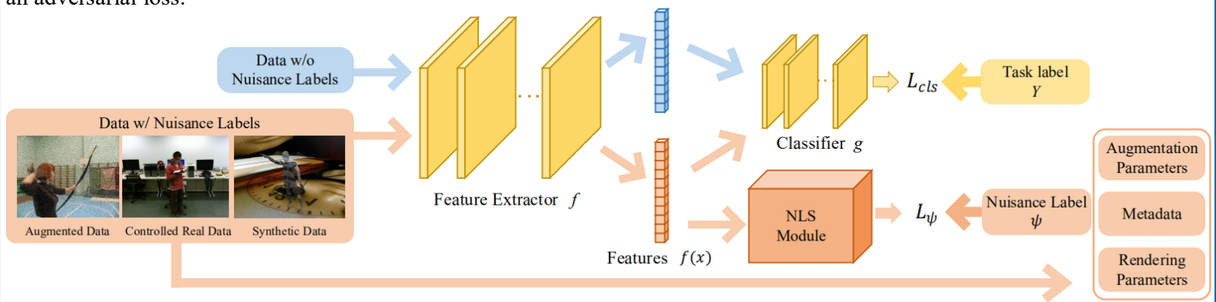
Nuisance Labels Collection

1. Data augmentation actually offers more information than people used to expect; the image processing parameters can be used in training as nuisance labels.
2. Some existing real datasets also contain such information, i.e. metadata during data collection.
3. Synthetic data generation offers a convenient way to manipulate multiple factors in the virtual world and we can easily reach the ground truth of these factors.



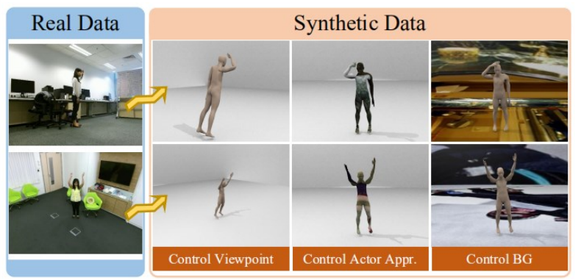
Pipeline overview

We utilize adversarial training to achieve the goal of "removing nuisance information". We divide input data into two parts: 1) data w/o nuisance labels, 2) data w/ nuisance labels. On the one hand, both inputs are put through a feature extractor and classifier the same as normal action recognition pipeline. On the other hand, data with nuisance labels is additionally put into NLS module for training feature representation. In this process, nuisance labels are used as extra supervision signals and output an adversarial loss.



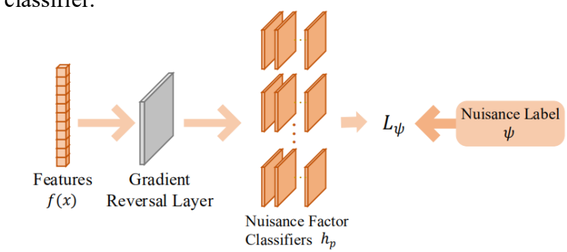
Datasets

1. MNIST-C with image corruption parameters.
2. UCF101 with image corruption parameters.
3. NTU RGB-D with real metadata.
4. NTU RGB-D with synthetic data parameters.



NLS module

The NLS module consists of two parts: gradient reversal layer and nuisance factor classifier. The nuisance factor classifier is to predict labels of a certain nuisance factor. The gradient reversal layer multiplies the gradient by a negative constant during the back-propagation process, making the input features as indistinguishable as possible for the nuisance factor classifier.



Results

Here we only show the results on NTU RGB-D, using NLS on real metadata and synthetic data parameters. More results are in the paper.

Model	Accuracy	F1 Score
Baseline	64.50	66.92
Real NLS	65.57	67.32
Sim Aug	68.75	70.99
Sim Aug+ANT1 × 1	68.04	71.05
Sim Aug+Sim NLS	70.40	73.03
Sim Aug+Real&Sim NLS	**70.64**	**73.38**

Table 3. I3D accuracy and F1 score on NTU RGB-D Cross Nuisance (CN) split. Real NLS refers to applying NLS to real metadata and Sim NLS refers to applying NLS to synthetic rendering parameters. Sim Aug refers to using CG synthetic data as data augmentation.

Method	Modality	CS	CV
Hands attention [2]	RGB+Skeleton	84.8	90.6
DA-Net [36]	RGB+Flow	88.1	92.0
Pose evolution [19]	RGB+Depth	91.7	95.3
Hands attention [2]	RGB	75.6	80.5
Pose evolution [19]	RGB	78.8	84.2
Multi-task [20]	RGB	85.5	-
Glimpse clouds [3]	RGB	86.6	93.2
I3D [4]	RGB	90.2	95.2
I3D + Real NLS	RGB	**90.7**	**95.6**

Table 4. Comparison with state-of-the-art on standard splits of NTU RGB-D dataset. Our method achieves the highest performance within all the approaches using only RGB input.