

Few-shot Action Recognition with Permutation-invariant Attention (Supplementary Material)

Hongguang Zhang^{1,2,3,5}, Li Zhang², Xiaojuan Qi^{4,2}, Hongdong Li^{1,5},
Philip H. S. Torr², Piotr Koniusz^{3,1}

¹ Australian National University, Canberra, Australia

² University of Oxford, Oxford, UK

³ Data61/CSIRO, Australia

⁴ The University of Hong Kong, Hong Kong, China

⁵ Australian Centre for Robotic Vision, Australia

1 Evaluation Protocols

Below we demonstrate the detailed training/validation/testing splits used in our paper. Though the limited number of previous works propose some evaluation splits on several action recognition datasets, they differ in every paper thus making it very difficult to produce fair comparisons with other works. Additionally, most of works use random train/test splits and have no validation set, thus making the results suffer from high variance and potentially overfitting to the test data. As we aim to fix these problems, we formally introduce three new evaluation protocols as the standard benchmarks, which can help compare models more accurately in a fair setting.

1.1 HMDB51

Actions of Train Split (31): brush hairs, catch, chew, clap, climb, climb stairs, dive, draw sword, dribble, drink, fall floor, flic flac, handstand, hug, jump, kiss, pullup, punch, push, ride bike, ride horse, shake hands, shoot bow, situp, stand, sword, sword exercices, throw, turn, walk, wave.

Actions of Validation Split (10): cartwheel, eat, golf, hit, laugh, shoot ball, shoot gun, smile, somersault, swing basketball.

Actions of Validation Split (10): fencing, kick, kick ball, pick, pour, pushup, run, sit, smoke, talk.

1.2 miniMIT

Actions of Train Split (120): arresting, assembling, attacking, baking, barbecuing, barking, bending, bicycling, biting, boating, bouncing, brushing, bulldozing, burning, camping, carrying, celebrating, chopping, clapping, cleaning, clinging, closing, combing, competing, covering, crawling, crying, cutting, descending, destroying, digging, dining, drawing, drenching, drilling, drinking, dripping, driving,

dropping, drying, dunking, emptying, entering, erupting, falling, filling, flipping, floating, flying, folding, frying, handwriting, hanging, hitting, juggling, kicking, knitting, landing, laughing, leaping, lecturing, lifting, mopping, opening, parading, photographing, picking, placing, pouring, pressing, protesting, pulling, pushing, rafting, raining, reading, removing, repairing, riding, rising, rowing, running, sawing, scratching, sewing, shaking, shaving, shopping, shouting, shredding, singing, skating, sleeping, slicing, sliding, smiling, smoking, snowing, speaking, spraying, spreading, sprinting, stacking, stirring, stitching, stretching, stroking, studying, swimming, swinging, tapping, tattooing, turning, twisting, typing, vacuuming, walking, washing, whistling, wrapping.

Actions of Validation Split (40): ascending, boiling, bubbling, chasing, combusting, constructing, cracking, crashing, crushing, diving, drumming, eating, exercising, gardening, grilling, grooming, hammering, hugging, inflating, licking, painting, peeling, pitching, planting, playing, playing sports, rolling, sanding, shoveling, smashing, spinning, steering, surfing, sweeping, tapping, throwing, unloading, watering, waving, wrestling.

Actions of Test Split (40): boxing, carving, catching, cheering, chewing, climbing, colliding, cooking, crafting, dancing, feeding, fishing, flooding, frowning, gripping, hiking, howling, jumping, launching, mowing, overflowing, pedaling, performing, piloting, playing music, racing, raising, resting, rubbing, sailing, slapping, sneezing, sniffing, splashing, storming, tying, waking, waxing, welding, yawning.

1.3 UCF101

Actions of Train Split (70): ApplyEyeMakeUp, Archery, BabyCrawling, BalanceBeam, BandMarching, BaseballPitch, Basketball, BasketballDunk, BenchPress, Biking, Billiards, BlowDryHair, BodyWeightSquats, Bowling, BoxingPunchingBag, BoxingSpeedBag, BreastStroke, BrushingTeeth, CricketBowling, Drumming, Fencing, FieldHockeyPenalty, FrisbeeCatch, FrontCrawl, Haircut, Hammering, HeadMassage, HulaHoop, JavelinThrow, JugglingBalls, JumpingJack, Kayaking, Knitting, LongJump, Lunges, MilitaryParade, Mixing, MoppingFlow, Nunchucks, ParallelBars, PizzaTossing, PlayingCello, PlayingDhol, PlayingFlute, PlayingPiano, PlayingSitar, PlayingTabla, PlayingViolin, PoleVault, Pullups, PushUps, Rafting, RopeClimbing, Rowing, ShavingBeard, Skijet, SoccerJuggling, SoccerPenalty, SumoWrestling, Swing, TableTennisShot, Taichi, ThrowDiscus, TrampolineJumpling, Typing, UnevenBars, WalkingWithDog, WallPushups, WritingOnBoard, YoYo.

Actions of Validation Split (10): ApplyLipstick, CricketShot, HammerThrow, HandstandPushups, HighJump, HorseRiding, PlayingDaf, PlayingGuitar, Shotput, SkateBoarding.

Actions of Test Split (21): BlowingCandles, CleanAndJerk, CliffDiving, CuttingInKitchen, Diving, FloorGymnastics, GolfSwing, HandstandWalking, HorseRace, IceDancing, JumpRope, PommelHorse, Punch, RockClimbingIndoor, SalsaSpin, Skiing, SkyDiving, StillRings, Surfing, TennisSwing, VolleyballSpiking.

2 Ablations on the Temporal Attention Unit

As our attention module is designed to be invariant to block permutations, one can ask whether such an invariance could be attained by simply processing one block at a time and discarding other neighbours. Thus, we investigate the performance by feeding separate temporal blocks into a 2D attention module to obtain the attention scores (in this experiment we neglect the long-term temporal consistency). On HMDB51 we obtain 41.68% and 56.03% for 1- and 5-shot protocols, respectively, which is worse than using our full temporal attention module (46.13% and 60.92%). This demonstrates that the attention unit can learn long-term dependencies even between permuted blocks when learning the temporal blocks of interests. This finding further supports our motivation for the self-supervised permutation-invariant attention alignment.