Localizing the Common Action Among a Few Videos - Supplementary File

Pengwan Yang[1,2*], Vincent Tao Hu[2*], Pascal Mettes[2], and Cees G. M. Snoek[2]

[1] Peking University, China
[2] University of Amsterdam, the Netherlands

1 Dataset

Detail of class statistics. We list the class names for the train, validation and test sets of the common instance and multi-instance ActivityNet in Table C, and for the common instance and multi-instance Thumos in Table D.

2 Method

Overview of the basic block and the residual block. The structure of the basic block is illustrated in Figure Aa. The main idea of the basic block is to align the features I_2 to the features I_1. Figure Ab depicts the schema of the residual block. Here, the residual block performs recalibration on the input features I.

3 Results

Numerical results of five-shot comparison. Table A displays the numerical results of the five-shot comparison between our method and modified versions of Buch *et al.* [i] and Hu *et al.* [ii] on both the common instance and multi-instance datasets. The numerical results are provided here to facilitate the comparison for the follow-up works.

    Effect of depth in the progressive alignment module. n in Equation 7 denotes the depth of the progressive alignment module which is set to 3 in practice. We ablate the effect of the depth of the progressive alignment module in Table B under the five-shot setting on the common instance ActivityNet dataset. From depth 1 to 3, the mAP improves steadily with only a slight increase in parameters. Results stagnate when the depth is higher than 3.

References

[i] Shyamal Buch, Victor Escorcia, Chuanqi Shen, Bernard Ghanem, and Juan Carlos Niebles. Sst: Single-stream temporal action proposals. In *CVPR*, 2017. 1
[ii] Tao Hu, Pascal Mettes, Jia-Hong Huang, and Cees GM Snoek. Silco: Show a few images, localize the common object. In *ICCV*, 2019. 1
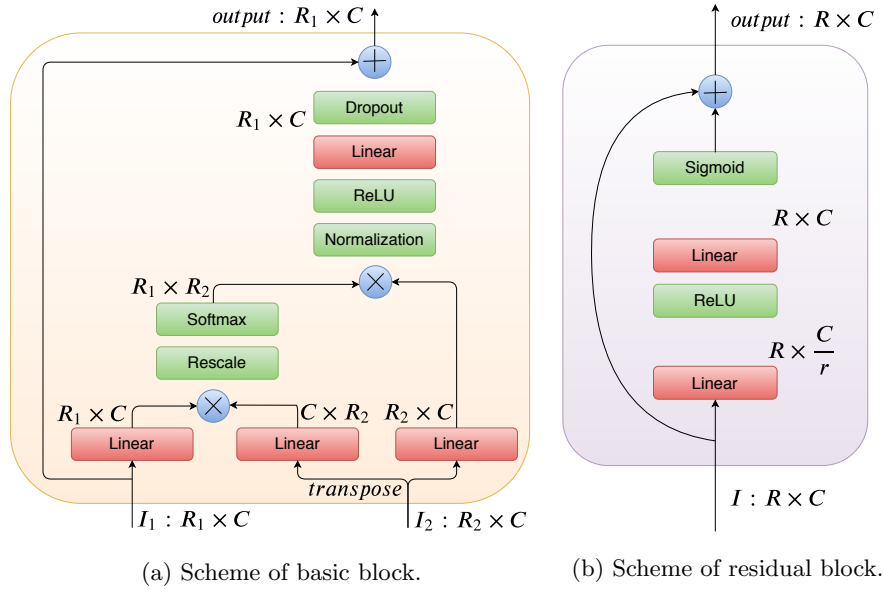
---

* Equal contribution.

(a) Scheme of basic block.   (b) Scheme of residual block.

Fig. A: **Overview of our blocks.** I_1, I_2 denote the inputs of our basic block (left) and I denotes the input of our residual block (right). $\otimes$ denotes matrix multiplication and $\oplus$ is element-wise sum. r is set to 4.

Table A: **Numerical results of five-shot comparison.**

	ActivityNet						Thumos					
	0.5	0.6	0.7	0.8	0.9	0.5:0.9	0.5	0.6	0.7	0.8	0.9	0.5:0.9
Common instance												
Buch *et al.*	39.7	33.6	27.0	14.0	4.6	23.3	35.7	29.4	20.8	11.7	3.4	20.2
Hu *et al.*	45.4	35.0	29.9	17.6	5.2	27.0	42.2	32.6	20.3	13.7	5.2	22.8
This paper	**56.5**	**47.0**	**37.4**	**21.5**	**11.9**	**34.9**	**51.9**	**42.7**	**24.4**	**17.7**	**10.1**	**29.3**
Common multi-instance												
Buch *et al.*	30.4	25.1	19.6	12.9	6.6	18.9	2.7	1.9	1.4	0.9	0.4	1.5
Hu *et al.*	38.9	27.2	18.3	12.7	7.3	20.9	6.8	3.1	2.2	1.8	1.3	3.1
This paper	**43.9**	**37.4**	**20.2**	**13.4**	**7.7**	**24.5**	**8.6**	**5.6**	**3.8**	**2.5**	**1.7**	**4.4**

Table B: **Ablation study on the depth of the progressive alignment module** under five-shot setting on common instance ActivityNet. From depth 1 to 3, the mAP improves steadily with a slight increase in parameters. Results stagnate when the depth is more than 3.

	n-**depth**				
	1	2	3	4	5
#param	47.2M	48.2M	49.3M	50.3M	51.4M
mAP	**53.9**	**55.6**	**56.5**	**56.7**	**56.4**

Table C: **Subset labels on common (multi-)instance ActivityNet.**

Training: Fun sliding down, Beer pong, Getting a piercing, Shoveling snow, Kneeling, Tumbling, Playing water polo, Washing dishes, Blowing leaves, Playing congas, Making a lemonade, Playing kickball, Removing ice from car, Playing racquetball, Swimming, Playing bagpipes, Painting, Assembling bicycle, Playing violin, Surfing, Making a sandwich, Welding, Hopscotch, Gargling mouthwash, Baking cookies, Braiding hair, Capoeira, Slacklining, Plastering, Changing car wheel, Chopping wood, Removing curlers, Horseback riding, Smoking hookah, Doing a powerbomb, Playing ten pins, Getting a haircut, Playing beach volleyball, Making a cake, Clean and jerk, Trimming branches or hedges, Drum corps, Windsurfing, Kite flying, Using parallel bars, Doing kickboxing, Cleaning shoes, Playing field hockey, Playing squash, Rollerblading, Playing drums, Playing rubik cube, Sharpening knives, Zumba, Raking leaves, Bathing dog, Tug of war, Ping-pong, Using the balance beam, Playing lacrosse, Scuba diving, Preparing pasta, Brushing teeth, Playing badminton, Mixing drinks, Discus throw, Playing ice hockey, Doing crunches, Wrapping presents, Hand washing clothes, Rock climbing, Cutting the grass, Wakeboarding, Futsal, Playing piano, Baton twirling, Mooping floor, Triple jump, Longboarding, Polishing shoes, Doing motocross, Arm wrestling, Doing fencing, Hammer throw, Shot put, Playing pool, Blow-drying hair, Cricket, Spinning, Running a marathon, Table soccer, Playing flauta, Ice fishing, Tai chi, Archery, Shaving, Using the monkey bar, Layup drill in basketball, Spread mulch, Skateboarding, Canoeing, Mowing the lawn, Beach soccer, Hanging wallpaper, Tango, Disc dog, Powerbocking, Getting a tattoo, Doing nails, Snowboarding, Putting on shoes, Clipping cat claws, Snow tubing, River tubing, Putting on makeup, Decorating the Christmas tree, Fixing bicycle, Hitting a pinata, High jump, Doing karate, Kayaking, Grooming dog, Bungee jumping, Washing hands, Painting fence, Doing step aerobics, Installing carpet, Playing saxophone, Long jump, Javelin throw, Playing accordion, Smoking a cigarette, Belly dance, Playing polo, Throwing darts, Roof shingle removal, Tennis serve with ball bouncing, Skiing, Peeling potatoes, Elliptical trainer, Building sandcastles, Drinking beer, Rock-paper-scissors, Using the pommel horse, Croquet, Laying tile, Cleaning windows, Fixing the roof, Springboard diving, Waterskiing, Using uneven bars, Having an ice cream, Sailing, Washing face, Knitting, Bullfighting, Applying sunscreen, Painting furniture, Grooming horse, Carving jack-o-lanterns

Validation: Swinging at the playground, Dodgeball, Ballet, Playing harmonica, Paintball, Cumbia, Rafting, Hula hoop, Cheerleading, Vacuuming floor, Playing blackjack, Waxing skis, Curling, Using the rowing machine, Ironing clothes, Playing guitarra, Sumo, Putting in contact lenses, Brushing hair, Volleyball

Testing: Hurling, Polishing forniture, BMX, Riding bumper cars, Starting a campfire, Walking the dog, Preparing salad, Plataform diving, Breakdancing, Camel ride, Hand car wash, Making an omelette, Shuffleboard, Calf roping, Shaving legs, Snatch, Cleaning sink, Rope skipping, Drinking coffee, Pole vault

Table D: **Subset labels on common (multi-)instance Thumos.**

Training: BaseballPitch, BasketballDunk, Billiards, CleanAndJerk, CliffDiving, CricketBowling, CricketShot, Diving, FrisbeeCatch, GolfSwing, HammerThrow, High-Jump, JavelinThrow, LongJump, PoleVault, Shotput
Validation: SoccerPenalty, TennisSwing
Testing: ThrowDiscus, VolleyballSpiking