

Interpretable Embedding for Ad-Hoc Video Search

Jiaxin Wu, Chong-Wah Ngo

Department of Computer Science, City University of Hong Kong

1. Examples of predicted concepts by dual-task model

Figure 1 shows two examples of concepts decoded by the proposed BCE loss. In the first example, all the top-10 concepts are mentioned in the captions. In the second example, the concepts not being mentioned, such as “hand”, “hair”, “holding”, are also relevant to the video clip indeed.

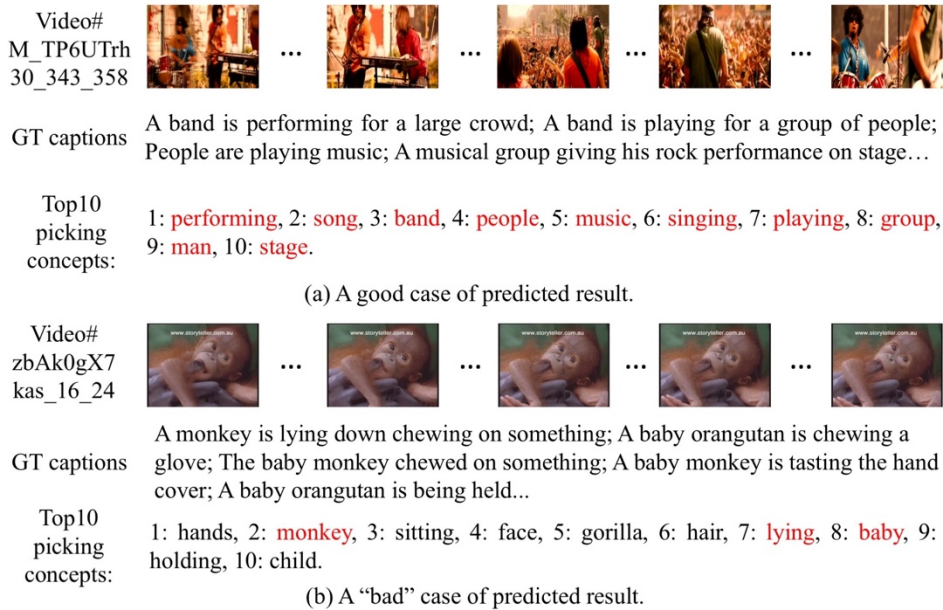


Figure 1. Ten highest activating concepts predicted based on the proposed class-sensitive BCE loss. The concepts marked in red are mentioned in the captions.

2. Text-to-video retrieval performances on video captioning datasets

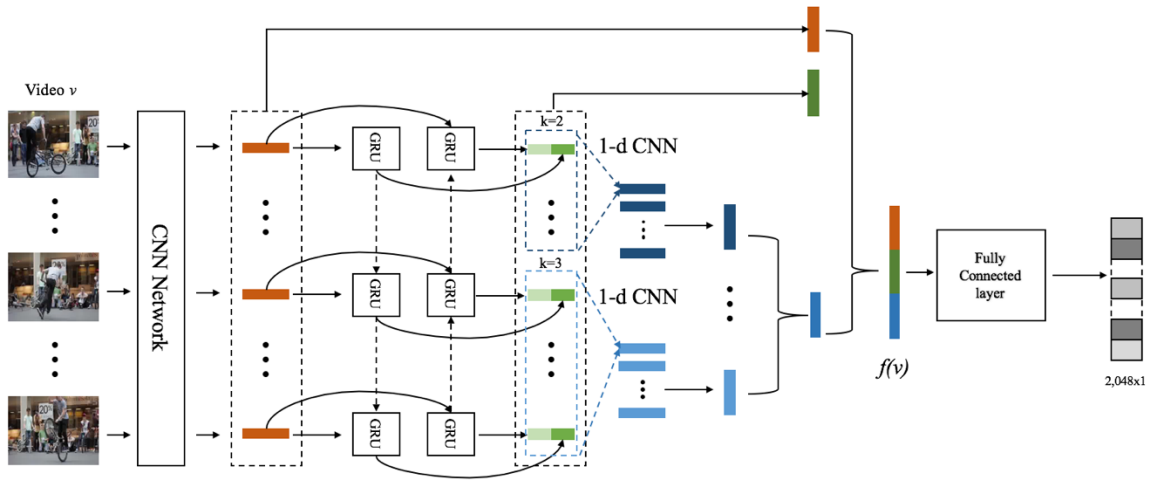
Table 1: Comparison with the Dual Coding [14] on MSR-VTT dataset based on ResNext101 video feature

Method	R@1	R@5	R@10	Med r	mAP
Dual Coding [14]	7.7	22.0	31.8	32.0	0.155
$DT_{combined}$	10.1	24.0	31.2	53.0	0.171

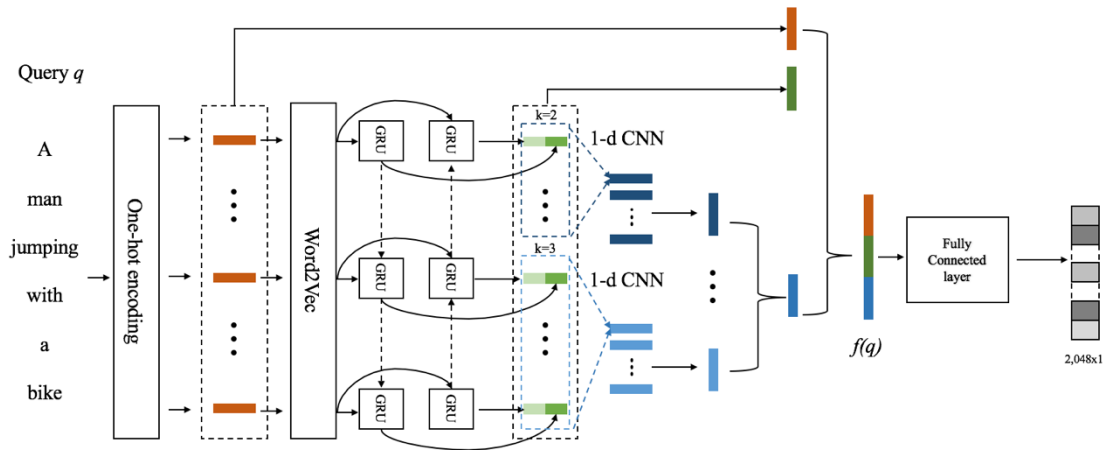
Table 2: Comparison with the Dual Coding [14] on MSVD dataset based on ResNext101-ResNet152 video feature

Method	R@1	R@5	R@10	Med r	mAP
Dual Coding [14]	18.5	42.2	53.9	8.0	0.299
$DT_{combined}$	25.1	45.7	54.8	7.0	0.350

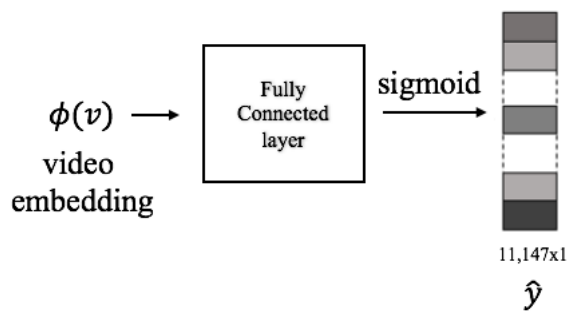
3. Details of the dual-task network



(a) Details of the visual encoding network



(b) Details of the textual encoding network



(c) Details of the visual decoding network

4. List of TRECvid query topics

tv16:

- 501 Find shots of a person playing guitar outdoors
- 502 Find shots of a man indoors looking at camera where a bookcase is behind him
- 503 Find shots of a person playing drums indoors
- 504 Find shots of a diver wearing diving suit and swimming under water
- 505 Find shots of a person holding a poster on the street at daytime
- 506 Find shots of the 43rd president George W. Bush sitting down talking with people indoors
- 507 Find shots of a choir or orchestra and conductor performing on stage
- 508 Find shots of one or more people walking or bicycling on a bridge during daytime
- 509 Find shots of a crowd demonstrating in a city street at night
- 510 Find shots of a sewing machine
- 511 Find shots of destroyed buildings
- 512 Find shots of palm trees
- 513 Find shots of military personnel interacting with protesters
- 514 Find shots of soldiers performing training or other military maneuvers
- 515 Find shots of a person jumping
- 516 Find shots of a man shake hands with a woman
- 517 Find shots of a policeman where a police car is visible
- 518 Find shots of one or more people at train station platform
- 519 Find shots of two or more men at a beach scene
- 520 Find shots of any type of fountains outdoors
- 521 Find shots of a man with beard talking or singing into a microphone
- 522 Find shots of a person sitting down with a laptop visible
- 523 Find shots of one or more people opening a door and exiting through it
- 524 Find shots of a man with beard and wearing white robe speaking and gesturing to camera
- 525 Find shots of a person holding a knife
- 526 Find shots of a woman wearing glasses
- 527 Find shots of a person drinking from a cup, mug, bottle, or other container
- 528 Find shots of a person wearing a helmet
- 529 Find shots of a person lighting a candle
- 530 Find shots of people shopping

tv17:

- 531 Find shots of one or more people eating food at a table indoors
- 532 Find shots of one or more people driving snowmobiles in the snow
- 533 Find shots of a man sitting down on a couch in a room
- 534 Find shots of a person talking behind a podium wearing a suit outdoors during daytime
- 535 Find shots of a person standing in front of a brick building or wall
- 536 Find shots of children playing in a playground
- 537 Find shots of one or more people swimming in a swimming pool
- 538 Find shots of a crowd of people attending a football game in a stadium

- 539 Find shots of an adult person running in a city street
- 540 Find shots of vegetables and/or fruits
- 541 Find shots of a newspaper
- 542 Find shots of at least two planes both visible
- 543 Find shots of a person communicating using sign language
- 544 Find shots of a child or group of children dancing
- 545 Find shots of people marching in a parade
- 546 Find shots of a male person falling down
- 547 Find shots of a person with a gun visible
- 548 Find shots of a chef or cook in a kitchen
- 549 Find shots of a blond female indoors
- 550 Find shots of a map indoors
- 551 Find shots of a person riding a horse including horse-drawn carts
- 552 Find shots of a person wearing any kind of hat
- 553 Find shots of a person talking on a cell phone
- 554 Find shots of a person holding or operating a tv or movie camera
- 555 Find shots of a person holding or opening a briefcase
- 556 Find shots of a person wearing a blue shirt
- 557 Find shots of person holding, throwing or playing with a balloon
- 558 Find shots of a person wearing a scarf
- 559 Find shots of a man and woman inside a car
- 560 Find shots of a person holding, opening, closing or handing over a box

tv18:

- 561 Find shots of exactly two men at a conference or meeting table talking in a room
- 562 Find shots of a person playing keyboard and singing indoors
- 563 Find shots of one or more people on a moving boat in the water
- 564 Find shots of a person in front of a blackboard talking or writing in a classroom
- 565 Find shots of people waving flags outdoors
- 566 Find shots of a dog playing outdoors
- 567 Find shots of people performing or dancing outdoors at nighttime
- 568 Find shots of one or more people hiking
- 569 Find shots of people standing in line outdoors
- 570 Find shots of a projection screen
- 571 Find shots of any type of Christmas decorations
- 572 Find shots of two or more cats both visible simultaneously
- 573 Find shots of medical personnel performing medical tasks
- 574 Find shots of two people fighting
- 575 Find shots of a person pouring liquid from one container to another
- 576 Find shots of a person holding his hand to his face
- 577 Find shots of two or more people wearing coats
- 578 Find shots of a person in front of or inside a garage
- 579 Find shots of one or more people in a balcony
- 580 Find shots of an elevator from the outside or inside view

- 581 Find shots of a person sitting on a wheelchair
- 582 Find shots of a person climbing an object (such as tree, stairs, barrier)
- 583 Find shots of a person holding, talking or blowing into a horn
- 584 Find shots of a person lying on a bed
- 585 Find shots of a person with a cigarette
- 586 Find shots of a truck standing still while a person is walking beside or in front of it
- 587 Find shots of a person looking out or through a window
- 588 Find shots of a person holding or attached to a rope
- 589 Find shots of car driving scenes in a rainy day
- 590 Find shots of a person where a gate is visible in the background

tv19:

- 611 Find shots of a drone flying
- 612 Find shots of a truck being driven in the daytime
- 613 Find shots of a door being opened by someone
- 614 Find shots of a woman riding or holding a bike outdoors
- 615 Find shots of a person smoking a cigarette outdoors
- 616 Find shots of a woman wearing a red dress outside in the daytime
- 617 Find shots of one or more picnic tables outdoors
- 618 Find shots of coral reef underwater
- 619 Find shots of one or more art pieces on a wall
- 620 Find shots of a person with a painted face or mask
- 621 Find shots of a person in front of a graffiti painted on a wall
- 622 Find shots of a person in a tent
- 623 Find shots of a person wearing shorts outdoors
- 624 Find shots of a person in front of a curtain indoors
- 625 Find shots of a person wearing a backpack
- 626 Find shots of a race car driver racing a car
- 627 Find shots of a person holding a tool and cutting something
- 628 Find shots of a man and a woman holding hands
- 629 Find shots of a black man singing
- 630 Find shots of a man and a woman hugging each other
- 631 Find shots of a man and a woman dancing together indoors
- 632 Find shots of a person running in the woods
- 633 Find shots of a group of people walking on the beach
- 634 Find shots of a woman and a little boy both visible during daytime
- 635 Find shots of a bald man
- 636 Find shots of a man and a baby both visible
- 637 Find shots of a shirtless man standing up or walking outdoors
- 638 Find shots of one or more birds in a tree
- 639 Find shots of inside views of a small airplane flying
- 640 Find shots of a red hat or cap

5. List of Boolean queries

“AND” queries:

B001 woman and man

B002 hand and head

B003 parachute and grass

B004 monitor and keyboard

B005 curtain and sofa

“OR” queries:

B006 holding or riding bike

B007 people sitting on chair or sofa

B008 walking or running

B009 skiing or skating

B010 football or basketball

“NOT” queries

B011 beach not man

B012 face not woman

B013 drinking not wine or beer

B014 flower not red or yellow

B015 two people kissing not bride and groom

6. AVS performances (xinfAP) of our dual-task models on each query

Table 3: AVS performances on query sets tv16-tv19

Query id	$DT_{concept}$	$DT_{embedding}$	$DT_{combined}$
501	0.123	0.244	0.308
502	0.171	0.376	0.385

503	0.277	0.356	0.369
504	0.140	0.167	0.166
505	0.150	0.208	0.242
506	0.464	0.206	0.318
507	0.676	0.765	0.756
508	0.033	0.044	0.056
509	0.351	0.165	0.277
510	0.373	0.267	0.337
511	0.019	0.024	0.026
512	0.325	0.167	0.250
513	0.003	0.004	0.003
514	0.079	0.191	0.181
515	0.009	0.017	0.021
516	0.011	0.020	0.023
517	0.153	0.129	0.148
518	0.285	0.245	0.266
519	0.107	0.222	0.246
520	0.001	0.000	0.000
521	0.073	0.097	0.113
522	0.047	0.210	0.215
523	0.013	0.044	0.036
524	0.026	0.115	0.125
525	0.019	0.045	0.040
526	0.142	0.130	0.163
527	0.056	0.074	0.104
528	0.258	0.177	0.219
529	0.009	0.010	0.011
530	0.061	0.165	0.141
Mean	0.148	0.163	0.185
531	0.134	0.127	0.123
532	0.062	0.060	0.112
533	0.253	0.223	0.228
534	0.051	0.123	0.117
535	0.049	0.016	0.021
536	0.032	0.044	0.046
537	0.160	0.567	0.528
538	0.465	0.241	0.316
539	0.000	0.016	0.012
540	0.356	0.284	0.338
541	0.450	0.451	0.450
542	0.391	0.181	0.257
543	0.000	0.501	0.265
544	0.023	0.042	0.053

545	0.063	0.080	0.080
546	0.000	0.003	0.001
547	0.001	0.113	0.104
548	0.486	0.377	0.529
549	0.082	0.163	0.196
550	0.006	0.237	0.193
551	0.225	0.673	0.652
552	0.048	0.492	0.481
553	0.111	0.152	0.160
554	0.015	0.054	0.062
555	0.043	0.101	0.097
556	0.048	0.198	0.216
557	0.145	0.193	0.209
558	0.470	0.182	0.394
559	0.192	1.024	0.959
560	0.050	0.037	0.041
Mean	0.147	0.232	0.241
561	0.037	0.035	0.037
562	0.032	0.120	0.088
563	0.171	0.498	0.506
564	0.225	0.236	0.247
565	0.519	0.307	0.407
566	0.019	0.056	0.065
567	0.002	0.008	0.008
568	0.021	0.038	0.037
569	0.014	0.013	0.013
570	0.000	0.002	0.002
571	0.022	0.123	0.094
572	0.152	0.055	0.086
573	0.063	0.103	0.133
574	0.172	0.035	0.040
575	0.070	0.078	0.075
576	0.000	0.161	0.138
577	0.490	0.316	0.395
578	0.008	0.036	0.021
579	0.015	0.005	0.008
580	0.003	0.034	0.018
581	0.358	0.303	0.356
582	0.002	0.059	0.031
583	0.022	0.210	0.173
584	0.145	0.297	0.289
585	0.129	0.295	0.291
586	0.003	0.031	0.046

587	0.003	0.022	0.023
588	0.017	0.027	0.027
589	0.009	0.038	0.035
590	0.004	0.013	0.014
Mean	0.091	0.118	0.123
611	0.100	0.184	0.220
612	0.003	0.083	0.072
613	0.019	0.012	0.014
614	0.043	0.069	0.107
615	0.006	0.228	0.218
616	0.003	0.034	0.033
617	0.020	0.003	0.006
618	0.351	0.323	0.356
619	0.002	0.067	0.047
620	0.289	0.445	0.442
621	0.509	0.511	0.524
622	0.137	0.156	0.162
623	0.233	0.298	0.368
624	0.101	0.068	0.113
625	0.293	0.243	0.286
626	0.195	0.181	0.213
627	0.011	0.049	0.038
628	0.005	0.184	0.194
629	0.013	0.084	0.092
630	0.051	0.101	0.107
631	0.010	0.281	0.263
632	0.001	0.047	0.046
633	0.105	0.274	0.307
634	0.001	0.036	0.036
635	0.234	0.374	0.378
636	0.029	0.252	0.303
637	0.257	0.160	0.240
638	0.165	0.051	0.081
639	0.004	0.001	0.002
640	0.263	0.232	0.278
Mean	0.115	0.168	0.185