Hierarchical Conditional Relation Networks for Video Question Answering

Thao Minh Le, Vuong Le, Svetha Venkatesh, Truyen Tran

Code: https://github.com/thaolmk54/hcrn-videoqa

Challenges of Video QA

- Understanding **temporal reasoning** in addition to **visual reasoning**.
- Videos are **richer** than images, can be incorporated with **additional channels** such as subtitle, speech etc.

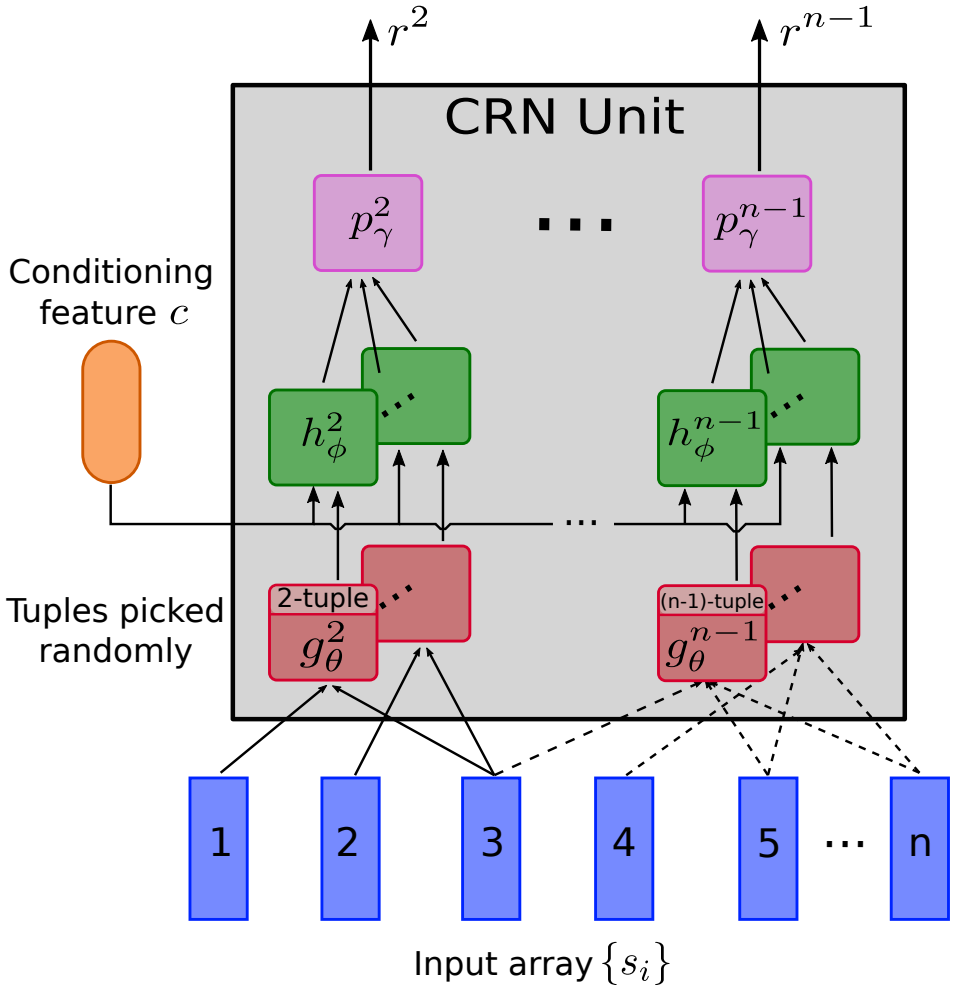
Q1: What does the boy with a brown hoodie do before running away ? ***A****: flip to the front side*

Q2: What does the boy with a brown hoodie do after flipping to the front side? ***A****: run away*

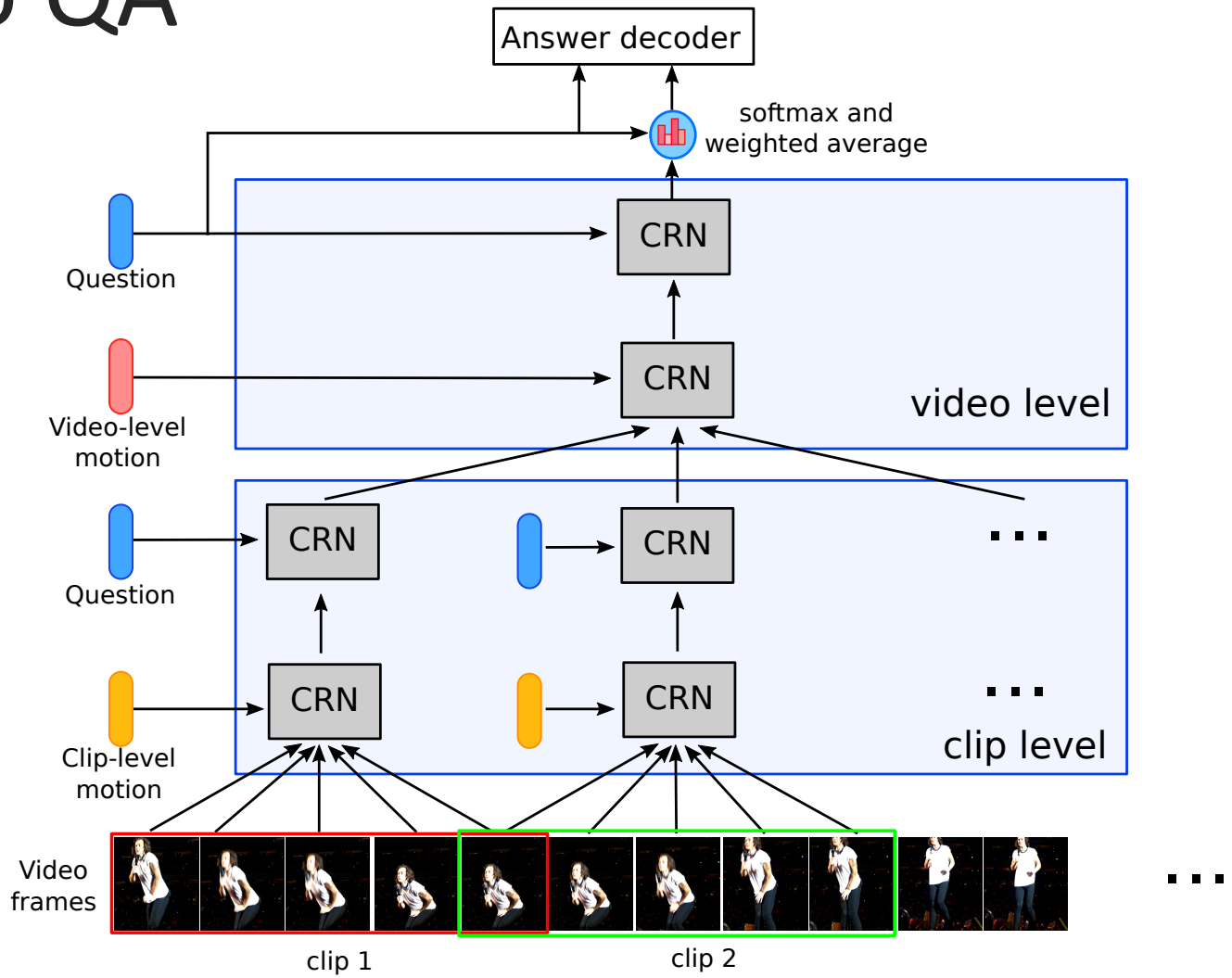
Q3: Where is the boy with brown jacket running? ***A****: street*

Conditional Relation Network Unit (CRN)

- A new **general-purpose neural unit** for representation and reasoning over videos.
- A **relational reasoning engine** operating on set with **cross modality conditioning**.



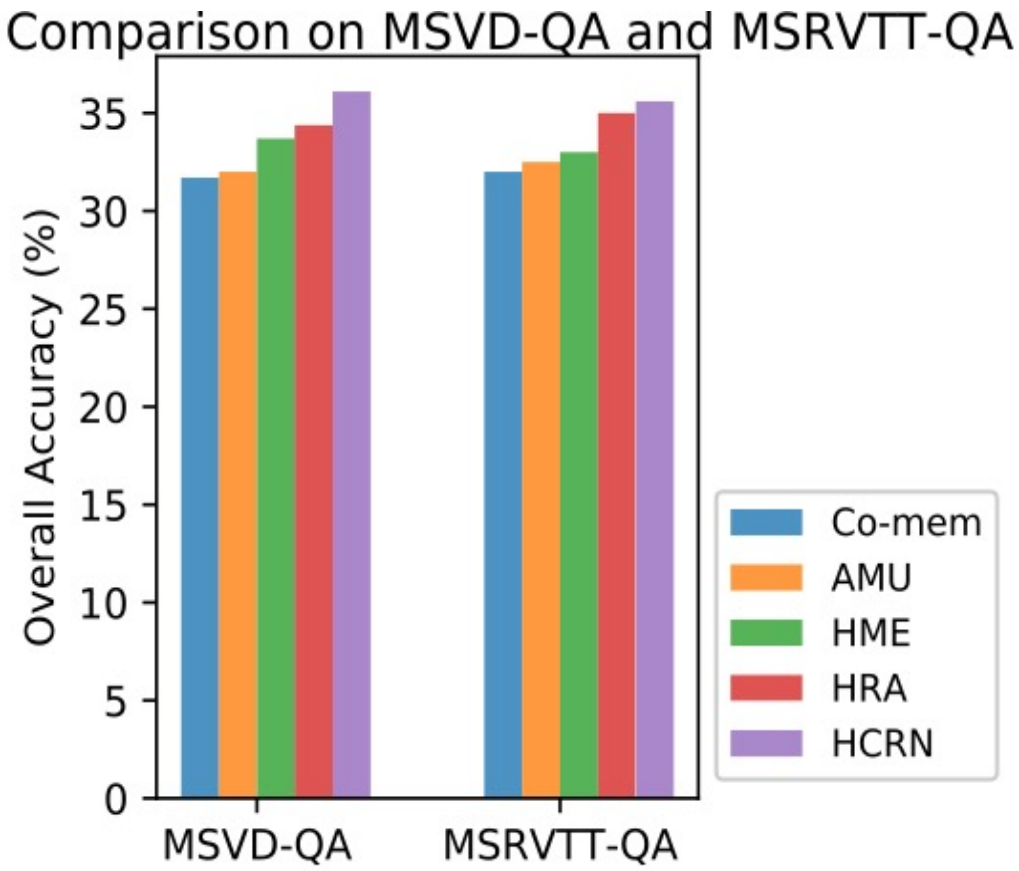
Hierarchical Conditional Relation Networks for Video QA



Results

Model	Action	Trans.	F.QA	Count
ST-TP	62.9	69.4	49.5	4.32
Co-Mem	68.2	74.3	51.5	4.10
PSAC	70.4	76.9	55.7	4.27
HME	73.9	77.8	53.8	4.02
HCRN	**75.0**	**81.4**	**55.9**	**3.82**

TGIF-QA dataset



THANK YOU!

Thao Minh Le

Email: lethao@deakin.edu.au

Personal Site: https://thaolmk54.github.io

Applied Artificial Intelligence Institute,

Deakin University,

75 Pigdons Rd, Waurn Ponds VIC, Australia, 3216