EVENT RECOGNITION IN HOCKEY WITH A MULTI-TASK LEARNING APPROACH

Pascale Walters
CS 886 - 002

April 15, 2020

Contents

- Introduction
- Related Work
- Methodology
  - *Dataset*
  - *Network Design*
- Results and Discussion

Introduction

- Video analytics of hockey games can be used to provide teams with an advantage over their competitors
  - *Influence coaching strategies and management decisions, increase fan engagement*
- To develop methods for understanding hockey games with deep learning, large quantities of annotated training data are required for fully-supervised methods. Collecting this data is very time consuming and can be very expensive.

Introduction

- Many broadcast networks provide commentary from hockey experts while the game is playing. Since the commentary provides a description of what is happening during the game, it can be thought of as weak supervision.
- ChirpNet: a method for performing semi-supervised action classification for hockey games.

Related Work

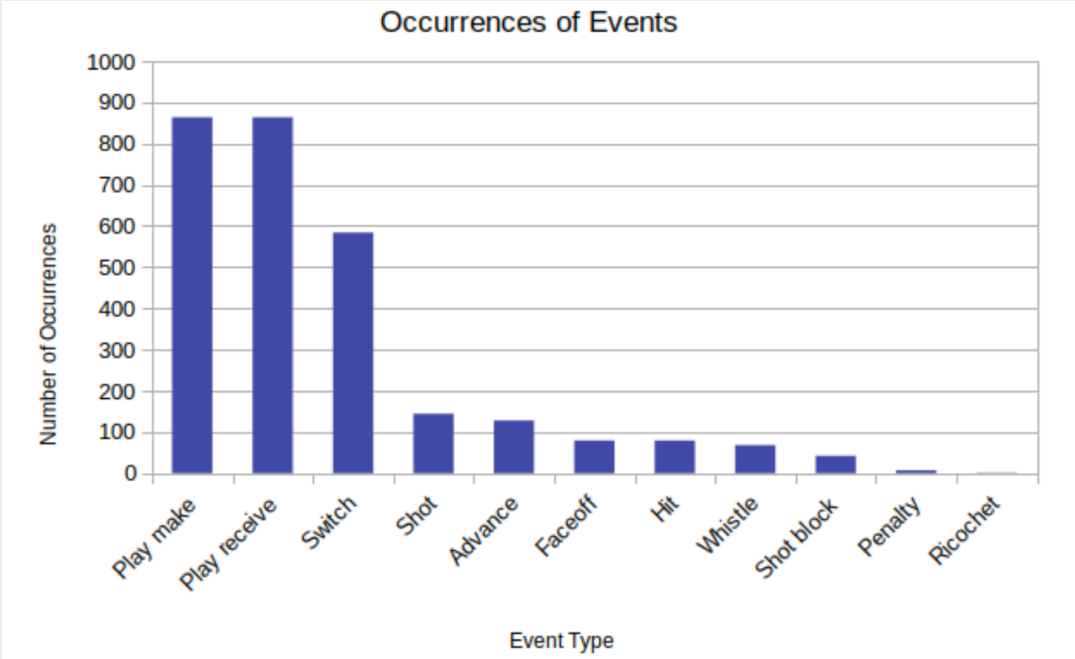
- Action recognition is an area of active research, including for sports games. There are several differences between general action recognition and that specifically for sports
  - *Varied camera viewpoints, camera motion (panning and zooming), rapid transitions between events*[1,2]
  - *Nuanced events, requires domain knowledge*
- In addition, there are fundamental differences between sports which makes a sports-agnostic method difficult
  - *Hockey: fast-paced, occlusions from boards, team interactions, small playing surface, line changes*

Related Work

- Tora *et al., 2017:* event classification in hockey using frames and player bounding boxes as input[1]
- Yu *et al.*, 2018: generate fine-grained video captions from basketball videos using visual input, model motion of players from each team and ball and their relationships, generate captions with two-layer bidirectional LSTM[3]
- Parmar *et al.*, 2019: action quality assessment for diving videos, use commentary and visual input to predict the score of dives[4]

Methodology: Dataset

- Annotations of event data were obtained from one NHL game at a one second resolution
  - *Event types: switch, advance, faceoff, play make, play receive, whistle, shot, shot block, hit, penalty, and ricochet*
- Divide game into short video clips that each contain ten events, then transcribe commentary with AWS Transcription service
  - *Train split: 198 clips, test split: 86 clips*
  - *Downsampled clips to 6 fps due to storage limitations*

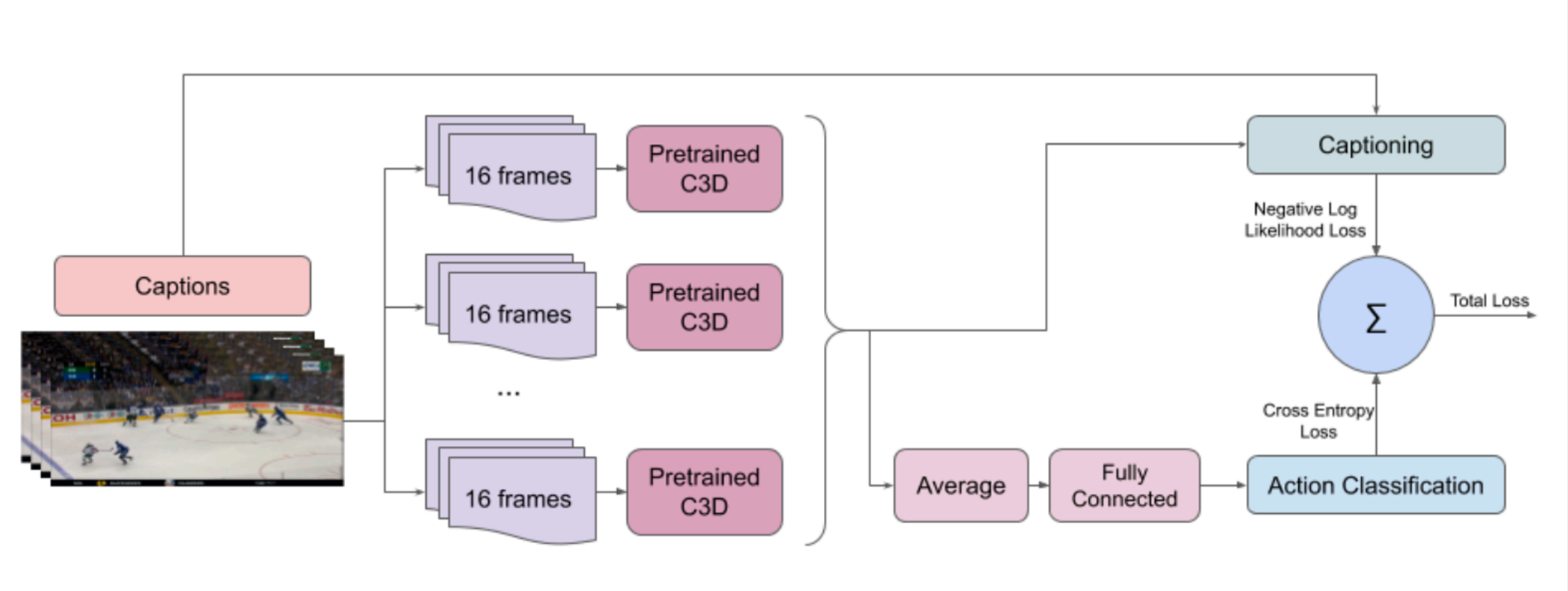


Methodology: Dataset

	**Events:** play make: 2, play receive: 3, switch: 1, shot: 1, faceoff: 2, whistle: 1 **Commentary:** Hands it over top. Quick, Quick goal here for Toronto throws a tour, and Zucker nearly got there before Hutchinson could cover. Let's watch it again. Just seven seconds after the opening face off. Well, it was a quick little bank off. It looked like it went off of suitor scape. Came back to martyr who cut back around. Spurgeon gets this backhand shot off, and Devon Dupnik did not look like he had stepped out to the top of his crease. There went for the butterfly, and it was a well placed shot. And he's a big goaltender, but probably needed to be out another foot. There. Toe. Get a piece of it. Great individual effort by martyr. My goodness.
	**Events:** play make: 3, play receive: 2, switch: 4, advance: 1 **Commentary:** trying to throw it towards the air trying for

Table 1: Example clips from the NHL dataset. Each clip has counts for all of the events, as well as a transcription of the commentary.

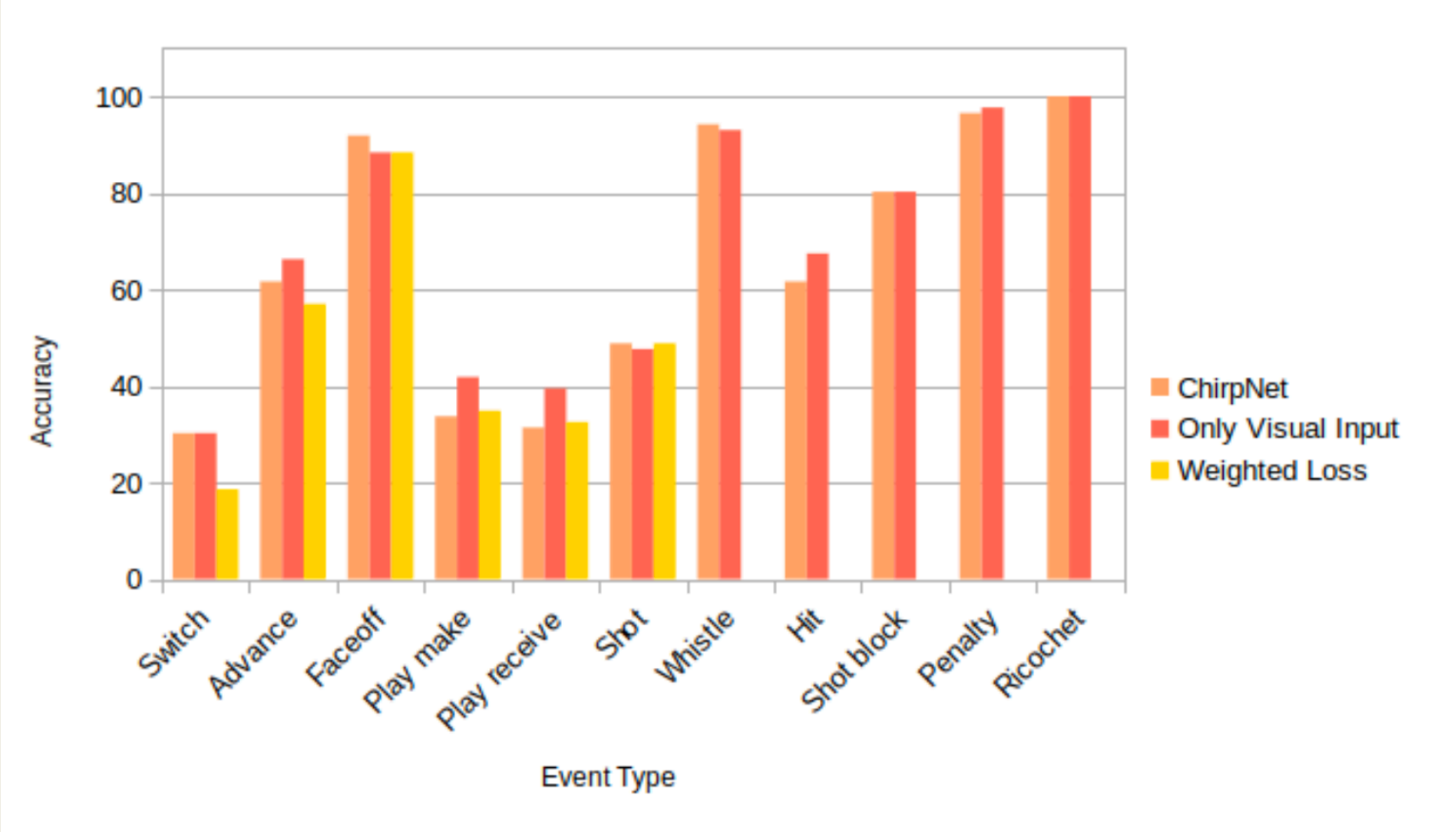
Methodology: ChirpNet



Results

Method	ChirpNet	Only Visual Input	Weighted Loss
Overall accuracy	66.39	68.39	-
Top class accuracy	49.61	52.32	46.71

Results


100
80
60
40
20
0
Accuracy
Switch
Advance
Faceoff
Play make
Play receive
Shot
Whistle
Hit
Shot block
Penalty
Ricochet
Event Type
ChirpNet
Only Visual Input
Weighted Loss

Discussion

- Including the captions allowed for ChirpNet to beat the baseline (visual input only) for some classes, but had worse overall performance
  - *Small dataset*
  - *Variable quality of annotations and transcription*
- ChirpNet assumes that all events are independent, which is likely untrue
- Future research:
  - *Larger dataset*
  - *Include player detections as input to the model*[1,3]
  - *Word sub-sequence kernel classifier for predicting whether the commentary refers to the events being shown*[2]

References

- [1] M. R. Tora, J. Chen, and J. J. Little, “Classification of puck possession events in ice hockey,” in *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pp. 147–154, IEEE, 2017.
- [2] S. Gupta and R. J. Mooney, “Using closed captions to train activity recognizers that improve video retrieval,” in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 30–37, IEEE, 2009.
- [3] H. Yu, S. Cheng, B. Ni, M. Wang, J. Zhang, and X. Yang, “Fine-grained video captioning for sports narrative,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6006–6015, 2018.
- [4] P. Parmar and B. T. Morris, “What and how well you performed? a multitask learning approach to action quality assessment,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 304–313, 2019.