ELEC97052/ELEC97053

Selected Topics in Machine Learning

Coursework on Recurrent Neural Networks [35% mark]

Figure 1: First-person hand action examples[1]

Release on 2 Mar, the report due on 29 Mar (midnight)

The coursework requires Python/Tensorflow (or other deep learning tool you prefer) programming. In all questions, you can use any existing toolbox/code, unless specified.

Submission instructions:

One joint report by each pair

Page limit: 4 A4 pages (**no appendix**) per report with 10 font size (use the IEEE standard double column paper format, either in MS word or latex).

http://www.pamitc.org/cvpr16/files/egpaper_for_review.pdf

http://www.pamitc.org/cvpr16/files/cvpr2016AuthorKit.zip

General principles for writing technical report are expected to be known and adhered to. Similarly for practices in conducting experiments, some are as listed below:

- **Unlike previous courseworks, appendix is not allowed. 4 page limit except for references.**
- Select relevant results that support the points you want to make rather than everything that the program code gives.
- The important results should be in the report.
- Use clear and tidy presentation style, consistent across the report e.g. figures, tables.
- The experiments should be described such that there is no ambiguity in the settings, protocol and metrics used.
- The main points are made clear, identifying the best and the worst case results or other important observations.
- Do not copy standard formulas from lecture notes, explain algorithms in detail, or copy figures from other sources. References to lecture slides or publications/webpages are enough in such cases, however short explanations of new terms or parameters referred to are needed.

Find and demonstrate the parameters that lead to optimal performance and validate it by presenting supporting results. Give insights, discussions, and reasons behind your answers.

Quality and completeness of discussions within the page limit will be marked. Include formulas where appropriate, results presented in figures, and their discussions. Try to visualise any interesting observations to support your answers.

Code required for the experiments can be taken from any public library if available, otherwise implemented if necessary.

Submit the report **in pdf** through the Blackboard system. No hard copy is needed. Write your full names and CID numbers on the first page.

If you have questions, please contact

Razvan Caramalau (r.caramalau18@imperial.ac.uk)
Rumeysa Bodur (r.bodur18@imperial.ac.uk)
Pedro Castro (p.castro18@imperial.ac.uk)

Train and test Recurrent Neural Networks (RNN) to classify first-person hand actions using 3D hand pose annotations. We provide a subset of the First-Person Hand Action Benchmark [1] with +1000 sequences of +50 actions. The hand pose information given to you consists of different sized sequences of the 3D location (normalized) of 21 hand keypoints. Additional pre-processing might be needed.

"The joints are ordered in this way: [Wrist, TMCP, IMCP, MMCP, RMCP, PMCP, TPIP, TDIP, TTIP, IPIP, IDIP, ITIP, MPIP, MDIP, MTIP, RPIP, RDIP, RTIP, PPIP, PDIP, PTIP], where 'T', 'I', 'M', 'R', 'P' denote 'Thumb', 'Index', 'Middle', 'Ring', 'Pinky' fingers. 'MCP', 'PIP', 'DIP', 'TIP' as in the following Figure:" [2]

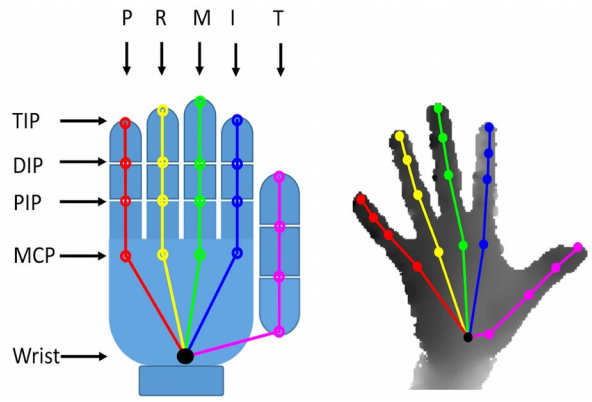

Figure 2: Hand annotation format. [2]

Q1. RNN Classifier [20 points] Design and describe an architecture for an RNN classifier. Train the RNN using the training set provided and test on the validation set. As input, your classifier should receive a sequence of 3D hand keypoints in the form of [N, 21, 3] and output the correspondent class, where N is the sequence length.
Please discuss, with results, each of the following topics:

- Steps to support multiple sized inputs
- Architecture choices (RNN/GRU/LSTM)
- Number of layers / hidden units
- Hyperparameters (learning rate / batch-size)
- Confusion matrix

Q2. Hierarchical RNN [15 points] The hand's structure can be used to your advantage. Fig. 2 provides an illustration of the data annotation. Update your Q1 classifier in order to support a hierarchical structure (page 27 of RNN slides) by **implementing intermediate finger representations**.

- Describe and justify your hierarchical structure
- Compare, measure and discuss the improvements between the naive (Q1) and hierarchical RNN approaches

Q3. Bonus [5 points] Push the limits of your model and data. Aim to achieve high accuracies (+90%). Larger capacity models are not the key.
Some suggestions:

- Data augmentation
- Ensemble
- Attention mechanism [3]
- Bidirectional mechanism [4]

[1] Garcia-Hernando, Guillermo et al. "First-Person Hand Action Benchmark with RGB-D Videos and 3D Hand Pose Annotations." CVPR 2018
[2] 5th International Workshop on Observing and Understanding Hands in Action, ICCV 2019
[3] Yan, Shiyang et al. "Hierarchical Multi-Scale Attention Networks for Action Recognition." Signal Processing: Image Communication 61 (2018)
[4] Xu, Kelvin et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:2048-2057, 2015.
[5] Wang, Cheng et al. "Image Captioning with Deep Bidirectional LSTMs." Proceedings of the 2016 ACM on Multimedia Conference