

---

**[2K5-ES-2]Machine learning: Multimedia**

Chair: Hiroki Shibata (Tokyo Metropolitan University)

Wed. Jun 10, 2020 3:50 PM - 5:30 PM Room K (jsai2020online-11)

---

4:30 PM - 4:50 PM

**[2K5-ES-2-03]Gated extra memory recurrent unit for learning video representations**○Daria Vazhenina<sup>1</sup>, Atsunori Kanemura<sup>1</sup> (1. Leapmind Inc.)

Keywords:Video representations, ConvRNNs, Video frame prediction

Convolutional recurrent neural networks (ConvRNNs) are widely used for spatiotemporal modelling tasks including video frame prediction. A major drawback of existing ConvRNNs is the amounts of computing and memory resources, which can hinder practical applications on embedded devices. Thus, to reduce them, we propose 1) a new gated architecture of the recurrent unit with temporal memory and 2) the replacement of computationally demanding convolution with more light-weight Hadamard product. Adopting such constraints can degrade the performance, but we show that the proposed model produces better results with reduced computation and memory. Quantitative evaluation with the Moving MNIST dataset shows that the overall performance of video frame prediction is improved by 13% in terms of MSE and by 3% in terms of SSIM without increasing the number of parameters and their multiplications, compared with the conventional ConvLSTM baseline. Further, applying the Hadamard product replacement outperforms the baseline MSE by 5%, while reducing the number of parameters by 14% and the number of multiplications by 25%.