

## **Animal Behavior Video Classification by Spatial LSTM**

**Huy Q. Nguyen<sup>1</sup>, Kasthuri Arachchi, S.P.<sup>2</sup>, Maduranga, M.W.P.<sup>3</sup> and Timothy K. Shih<sup>4</sup>**

<sup>1,2,4</sup>Dept. of CSIE, and Information Engineering, National Central University, Taiwan

<sup>3</sup>Dept. of Electrical and Electronic Technology, University of Vocational Technology, Sri Lanka

Email: {huyng121988, sandelik}@gmail.com, pasan@univotec.ac.lk, timothykshih@gmail.com

### **Abstract**

Deep learning which is basis for building artificial intelligent system is become a quite hot research area in recent years. Current deep neural network increase human recognition level of natural images even with huge dataset such as ImageNet. Among successful architectures, Convolution Neural Network (CNN) and Long Short-term Memory (LSTM) are widely used to build complex models because of their advantages. CNN reduces number of parameters compare to full connected neural net. Furthermore, it learns spatial features by sharing weights between convolution patch, which is not only help to improve performance but also extract similar features of input. LSTM is an improvement of Vanilla Recurrent Network (RNN). When processing with time-series data, RNN gradient has tend to vanish in training with backpropagation through time (BTT), while LSTM proposed to solve vanish problem. Therefore it is well suited for manage long-term dependencies. In other words, LSTM learn temporal features of time-series data. During this we study focused on creating an animal video dataset and investigating the way that deep learning system learn features with animal video dataset. We proposed a new dataset and experiments using two types of spatial-temporal LSTM, which allow us, discover latent information of animal videos. According to our knowledge of previous studies, no one has used this method before with animal activities. Our animal dataset created under three conditions; data must be videos. Thus, our network can learn spatial-temporal features, objects are popular animals like cats and dogs since it is easy to collect more data of them and the third is one video should have one animal but without humans or any other moving objects. Under experiments, we did the recognition task on Animal Behavior Dataset with two types of models to investigate its' differences. The first model is Conv-LSTM which is an extend version of LSTM, by replacing all input and output connections of convolutional connections. The second model is Long-term Recurrent Convolutional Networks (LRCN), which proposed by Jeff Donahue. More layers of LSTM units can easily added to both models in order to make a deeper network. We did classification using 900 training and 90 testing videos and could reached the accuracy of 66.7% on recognition rate. Here we did not do any data augmentation. However in the future we hope to improve our accuracy rate using some of preprocessing steps such as flip, rotate video clips and collecting more data for the dataset.

**Keywords:** *Spatial LSTM, LRCN, Animal behavior dataset, Video classification.*