



*“...real-world applications could be in video surveillance, or in human-robot interaction, to understand what is going on just in front of the robot e.g. what is the person doing in real time?”*

Fabien Baradel is a PhD student at INSA Lyon in France. He speaks to us about his poster presentation today. The co-author of the paper are Christian Wolf, an associate professor at INSA-Lyon and LIRS/CITI, Julien Mille, associate professor at INSA Centre Val de Loire and LI Tours, and Graham Taylor, associate professor at University of Guelph. The research project is funded by the ANR/NSERC DeepVision.

Fabien says his work is about human activity recognition, having a model

which is able to infer what is going on in a video using only the RGB data. Its real-world application could be in video surveillance, or in human-robot interaction, to understand what is going on just in front of the robot e.g. what is the person doing in real time?

In terms of human activity understanding, he says the real problem right now is that they are dealing with fine-grained understanding. There are some really local parts which are important and discriminative for the task, and right now, all the video understanding problems are more about understanding what’s going on in the context. This work is more about understanding a local part of the video.

He adds that most of the related work on this subject is more about using pose data information, so skeleton data which is given by the software, but in their case, they wanted to have a model which is able to run using only RGB data for understanding some fine-grained action done by humans.

*“We wanted our method to be fully differentiable...”*

Fabien explains further: “We wanted our method to be fully differentiable, to use backpropagation in order to learn our ways, so we have used a spatial transformer to do some cropping in the video in a differentiable way. This extracts a patch in the video, because we believe that extracting local information is important for the whole task.”

He tells us that they have been focusing on trained video and that is already a difficult enough task, but in terms of next steps, the biggest challenge – and one which would have a lot of applications in the real world –

would be to do an extension of this work on untrained video, so understanding when there is no action happening in the video and when the action starts and ends.

If you want more information about Fabien’s work you can visit the project page and there is also a github repository where they will be releasing the code this week and you will be able to train and evaluate their work.

You can also learn more and ask Fabien any questions about this work by coming along to his poster [H2] today at 10:10-12:30 in Halls C-E.

*“...in terms of next steps, the biggest challenge – and one which would have a lot of applications in the real world – would be to do an extension of this work on untrained video.”*

