RE-EXAMINING DISTILLATION FOR CONTINUAL OBJECT DETECTION

Eli Verwimp\textsuperscript{1} - Kuo Yang\textsuperscript{2} - Sarah Pariset\textsuperscript{2} - Lamqng Hong\textsuperscript{2} - Steven McDonagh\textsuperscript{2} - Eduardo Pérez Pelli\textsuperscript{2} - Matthias De Lange\textsuperscript{1} - Tinne Tuytelaars\textsuperscript{1}

\textsuperscript{1} KU Leuven - \textsuperscript{2}Huawei Noah's Ark Lab

\textbf{Introduction}

\textbf{Teacher}

\textbf{Student}

- Distillation losses
- Successful
- Unsuccessful
- Classification & regression losses

\textbf{2 GOALS}

Retain old knowledge from former tasks present in the teacher model, by using knowledge distillation from the teacher to the student model.

Learn new tasks from newly annotated data as well as possible. These can be new classes or new domains.

\textbf{Problems}

Not all classes from the new task are learned well

Badly trained!

New objects (e.g., person) were present in old task's images (e.g., horse). The teacher has learned to classify these as background or another class and these predictions are used in knowledge distillation. This is a problem when overlap is high (A) and lower (B).

\textbf{Solutions}

\textbf{Selective Distillation (solves A)}

Don't use proposals overlapping with new GT in distillation loss

\textbf{Huber Loss (solves B)}

Penalizes large differences less than MSE

\textbf{Results}

\textbf{Conclusion}

1. Most continual learning issues are in the ROI-heads
2. Careful application of distillation improves effectiveness
3. Room for improvement by leveraging rehearsal techniques (see paper)