Watch, Ask, Learn, and Improve: a lifelong learning cycle for visual recognition

Christoph Käding, Erik Rodner, Alexander Freytag, and Joachim Denzler
Computer Vision Group, Friedrich Schiller University Jena, Germany

The WALI System for learning animal classifiers from YouTube data

- **Watch**: download and watch YouTube videos autonomously
- **Ask**: actively select frames and ask human oracle for annotation
- **Learn**: incorporate new knowledge incrementally
- **Improve**: the knowledge grows over time

**Watch**
- download and watch “animal documentary” videos from YouTube
- build long term memory out of 500 images which occurred during the last 5 hours of video
- normalized relu7 features of BLVC AlexNet CNN [2] are used as feature representation
- every 10th frame is considered as unlabeled data

**Ask**
- one-vs-all classifier \( w_k^T x \) for each class \( k \in Y^d \)
- active selection with 1-vs-2 strategy [3]:
  \[
  k = \arg\max_{k \in Y^d \setminus \{k\}} w_k^T x \\
  q(x) = w_k^T x - \arg\max_{k \in Y^d \setminus \{k\}} w_k^T x
  \]
- avoid inappropriate images via reject strategy [4]:
  \[
  \tilde{q}(x) = (1 - p(\text{rejection} | x)) \cdot q(x)
  \]

**Learn**
- update classifiers incrementally
- linear regression with quadratic loss function yields possibility for efficient updates [5]

**Improve**
- active learning and discovery
- incremental learning

---