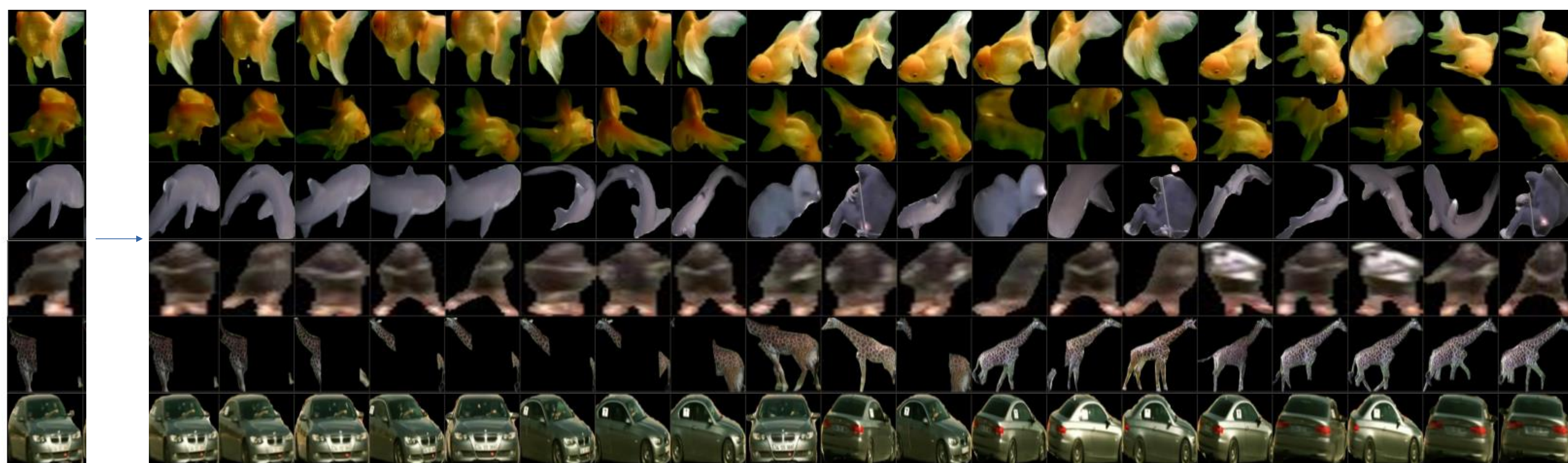




Video Instance Search for Criminal Investigation

An Thi Nguyen – an.t.nguyen@ntnu.no

Given an image of a specific object, we want to find occurrences of this object in a large video collection.



Query Search Results

Multi-modal
representation learning
from synthetic data is
used to align multi-view
representations with a
pre-aligned image-text
embedding space.

