Large-Scale Visual Recognition with Explicit Embedding

Florent Perronnin
Xerox Research Center Europe, France

Abstract

This lecture addresses Large-Scale Visual Recognition (LSVR), the problem of understanding visual content (e.g. photos or videos) on a large-scale. This is a topic which has received much attention in the computer vision community in the last few years: as larger datasets such as ImageNet (www.image-net.org) have become available, handling millions of images and thousands of label classes has become the norm rather than the exception. Since LSVR is a vast topic, I will mainly focus on the notion of explicit embedding. In the first part, I will review explicit input embedding (also referred to as explicit feature mapping) techniques which provide efficient learning and inference mechanisms when the number of training samples is large. In the second part, I will review explicit output embedding (also referred to as output encoding) techniques which address the problem of handling a large number of classes.

Keywords

large-scale learning, input embedding, output coding