
Large-Scale Historical Watermark Recognition: dataset and a new consistency-based approach

Mathieu Aubry^{*1}, Xi Shen , Oumayma Bounou , Tom Monnier , Ilaria Pastrolin ,
Christine Benevent , Marie-Françoise Limon-Bonnet , François Bougard , Marc H.
Smith , Olivier Poncet , and Pierre-Guillaume Raverdy

¹Laboratoire d'Informatique Gaspard-Monge – Ecole des Ponts ParisTech – France

Abstract

The study of watermarks is a key step for archivists and historians as it enables them to reveal the origin of paper. With a large number of well-defined classes, cluttered and noisy samples, different types of representations, both subtle differences between classes and high intra-class variation, historical watermarks are also challenging for pattern recognition. We recently introduced a new approach for this specific task which showed promising results. Overcoming the difficulty of data collection, we introduced a large public dataset with more than 6k new photographs, allowing for the first time to tackle at scale the scenarios of practical interest for scholars: one-shot instance recognition and cross-domain one-shot instance recognition amongst more than 16k fine-grained classes. We demonstrate that this new dataset is large enough to train modern deep learning approaches, and show that standard methods can be improved considerably by using mid-level deep features. Building upon this approach, this we proposed a new public web application dedicated to automatic watermark recognition entitled 'Filigranes pour tous'. The application not only hosts a detailed catalog of more than 17k watermarks manually collected from the French National Archives (Minutier central) or extracted from existing online resources (Briquet database), but it also enables non-specialists to identify a watermark from a simple photograph in a few seconds. Moreover, additional watermarks can easily be added by the users making the enrichment of the existing catalog possible through crowdsourcing.

Keywords: Watermarks, deep learning

^{*}Speaker