International Journal on Document Analysis and Recognition (IJDAR) (2023) 26:211–222
https://doi.org/10.1007/s10032-023-00439-6

ORIGINAL PAPER

Scheme for palimpsests reconstruction using synthesized dataset

Boraq Madi[1] · **Reem Alaasam**[1] · **Raed Shammas**[1] · **Jihad El-Sana**[1]

Received: 15 November 2022 / Revised: 19 March 2023 / Accepted: 26 April 2023 / Published online: 29 May 2023

Abstract
This paper presents **P**alimpsest **M**anuscripts **R**econstruction **G**enerative **A**dversarial **N**etwork, a novel framework for restoring the *original forms* of input palimpsests; it removes the over-text and refills the missing gaps in the text and background in an end-to-end manner. The structure and attributes of the under-text are encoded using reference patches. The generator network combines the encoded reference with an input palimpsest patch and restores the original form. To train our model, we synthesize palimpsests that mimic the attributes of the original ones. We compare the performance of our model with the state-of-art models using five different evaluation metrics, such as PSNR and SSIM. We show that our approach not only achieves state-of-the-art performance in terms of PSNR/SSIM metrics but also significantly improves the visual quality of the restored images.

Keywords Palimpsest · Occlusions · GANS · Arabic documents · Hebrew documents

1 Introduction

Prior to the invention of paper, untanned skins of animals, mainly sheep, goats, and cows, were used as writing material. This kind of skin is called Parchment and has been used as a writing medium for over two millennia. Due to the lack of parchment, many were scraped or washed off to be reused for other writing. The scraping is usually not perfect, and it is often possible to notice the under-text, which may have valuable historical information [1, 2].

Parchments that include (at least) two layers of text are denoted *palimpsests*. The main objective of studying palimpsest is recovering the original under-text. This task is challenging due to color degradation and complex over-text shape, which make deciphering palimpsests very difficult [3].

Collections of valuable palimpsests have been located in historical archives worldwide, and most have been digitized [4]. Archimedes Palimpsest is among the most famous of them. It is originally a Byzantine Greek copy of a publication of Archimedes and others, which was overwritten with a religious text.

The reconstruction of the underline text is time-consuming and expensive, as it often requires highly trained experts. Automatic reconsecration has attracted the interest of researchers, and an interesting body of work has been developed [1, 5–7]. However, some of these approaches require heavy pre-processing [7], and others assume the existence of prior knowledge concerning the text language and handwriting style [5] or require the existence of training images for the under-text, which is not always the case [5]. Most approaches aim at reconstructing the under-text or the over-text, and very few focus on reconstructing the *original form*, defined as the document before washing off the under-text and adding the over-text.

In this paper, we propose a novel approach to restore a palimpsest manuscript to its *original form* before washing off the under-text and adding the over-text, i.e., we remove the over-text from the input palimpsest, complete the broken characters of the under-text, and reconstruct the original background in one shot (similar to inpainting). Typical inpainting approaches manage to restore natural images well, but they often fail to reconstruct text accurately due to the critical semantic meaning of fine features. This necessitates the development of sophisticated approaches that not only

✉ Boraq Madi
borak@post.bgu.ac.il

Reem Alaasam
rym@post.bgu.ac.il

Raed Shammas
rshammes@post.bgu.ac.il

Jihad El-Sana
el-sana@cs.bgu.ac.il

[1] Computer Science, Ben-Gurion University of the Negev, Beersheba, Israel