

Determining the Influence of Papyrus Characteristics and Data Argumentation on Fragments Retrieval with Deep Metric Learning

In der Praxis sind altertümliche Papyri in mehrer Fragmente zerrissen. Die Aufgabe von Papyrologen ist es, diese Fragmente zusammenzusetzen und zu entziffern. Zusammen mit Historikern können dadurch wichtige Ergebnisse über das Altertum gewonnen werden. Allerdings ist der Prozess des händischen Zusammensetzens sehr zeitaufwendig und schwierig, weil Fragmente sich in Farbe, Struktur und Form so sehr unterscheiden, dass sie nicht perfekt zueinander passen - wie bei einem entworfenen Puzzle.

In der folgenden Arbeit geht es darum, Papyrologen die Arbeit zu erleichtern, indem der Prozess des Zusammensetzens teilautomatisiert wird. Hierfür wird ein Algorithmus entworfen, der zu einem Papyrus-Fragment eine kleinere Auswahl an mit hoher Wahrscheinlichkeit passenden Fragmenten vorschlägt. Zusammengefasst wird dieser Algorithmus im Folgenden eine Puzzlehilfe-Funktion genannt. Mit solch einer Puzzlehilfe-Funktion kann die Effektivität und Effizienz beim Zusammensetzen von Fragmenten erhöht werden. Allerdings muss eine Puzzlehilfe Funktion dabei mehrere Bedingungen erfüllen.

Zum einen darf das Erstellen einer solchen Funktion keine großen Datenmengen an zusammengesetzten Papyri Fragmenten voraussetzen, denn es existieren weltweit nur wenige solcher Daten. Deep Metric Learning (DML) wurde bisher verwendet, um eine Puzzlehilfe-Funktion zu erstellen und erzielt dabei einen top-1 Genauigkeitsgrad von 0.73. Der Top-1 Genauigkeitsgrad ist die Treffergenauigkeit Puzzlehilfe-Funktion, wobei jeweils nur der Kandidat mit der höchsten Wahrscheinlichkeit berücksichtigt wird. Wenn man das DML weiter verfeinern möchte, um eine sehr gute Genauigkeit zu erreichen, werden immens viele Daten benötigt. Andernfalls würde das Model sehr wahrscheinlich nicht konvergieren. Wenn die Fragmente händisch zusammensetzen werden, um somit einen großen Datensatz zu erhalten, wird der Zweck der Puzzlehilfe Funktion obsolet. Der Zeitaufwand für den Papyrologen bleibt im besten Fall derselbe oder wird im schlechtesten Fall erhöht. Generative-Adversarial-Networks (GANs) sind eine gut erforschte Möglichkeit, domänenübergreifend mehr künstliche Daten zu erzeugen. Es wurde gezeigt, dass die Genauigkeit über verschiedene Anwendungsbereiche hinweg dadurch signifikant erhöht werden kann. In der Arbeit soll ebenfalls ein GAN verwendet werden, um die das Problem der zu wenigen Daten zu umgehen.

Eine andere Bedingung ist, dass die falsch-negativen Treffer möglichst gering sind. In der Vorauswahl der Puzzlehilfe Funktion sollen also möglichst alle potenziell passenden Fragmente enthalten sein. Ist das passende Fragment nicht in der Teilmenge enthalten, hat der Papyrologe keine Chance, das richtige Fragment zu identifizieren. Es sei denn er weitet seine Suche wieder auf den gesamten Datensatz aus, was die Puzzlehilfe Funktion wieder obsolet werden lässt. Eine weiterer Nachteil ist, dass er möglicherweise extrem lange nach dem passenden Fragment sucht, welches aber nicht zu finden ist. Zielführende Metriken, um diese Eigenschaft zu Messen und ein Gesamtbild über das Model zu gewinnen, sind die mAP, top-1, pr@10, pr@100. Im Folgenden wird vereinfacht nur über Genauigkeit gesprochen. Diese Genauigkeit hängt stark davon ab, welcher Bildausschnitt als Eingabe verwendet wird. Genauer, was dem Model beim Trainieren als gleich und ungleich suggeriert wird, determiniert die Genauigkeit des Modells. In dieser Arbeit soll statistisch untersuchen, wie bestimmte Papyruscharakteristiken die Genauigkeit eines solchen Models beeinflussen. Hierfür wird die Genauigkeit verschiedener Modelle

verglichen, indem nur der Vordergrund eines Papyrus (Text) oder der Hintergrund (Fasern) als Eingabe zum trainieren verwendet wird. Das separieren von Text und Fasern ist ebenfalls eine Herausforderung, weil der Untergrund auf dem das Papyrus liegt oft die ähnliche Farben hat wie die Schrift auf dem Papyrus selbst. Wir verwenden verschiedene Methoden und evaluieren diese an der Gesamtgenauigkeit. Eine separierte Evaluierung wäre ohne eine Pixel genaue Ground Truth nicht möglich. Außerdem liegt der Fokus dieser Arbeit auf dem Einfluss der Charakteristiken auf die Genauigkeit und nicht auf dem Separieren selbst. Für die Texterkennung verwendet wird ein Method verwendet die auf einem Threshold basiert (Binarization), welcher den Text herauslöst und eine Text-Maske generiert. Die herausgelösten Stellen können nun mittels eines Inpainting-Verfahrens und der generierten Maske aufgefüllt werden. Auch das Inpainting-Verfahren wird nur anhand der Gesamtgenauigkeit evaluiert.

Zuletzt wird untersucht, ob bestimmte Charakteristiken (z.B. Fasern) noch auf andere weise verwendet werden können. Zum Beispiel, um den Papyrologen zu helfen, nicht nur potenzielle Kandidaten zu identifizieren, sondern auch deren genaue Position zu bestimmten. Dies wäre der logische Schritt hin zu einer vollautomatisierten Funktion. Das finale Ziel ist es, durch die Kombination des DML-Modells und einer geeigneten Methode zur Positionsbestimmung, eine Puzzler-Löser-Funktion zu entwerfen.

Zusammenfassend gliedert sich die Arbeit in folgende Meilensteine:

1. Separieren von Text und Papyrus-Fasern durch Binarization und Inpainting.
2. Erzeugen von größeren Datenmengen durch ein GAN.
3. Evaluierung mittels eines DML Models anhand der originalen Daten im originalen Zustand (Text und Fasern), im bearbeiteten Zustand (nur Text oder nur Fasern). Außerdem eine Evaluierung der künstlichen Daten zusammen mit den originalen Daten im originalen Zustand und bearbeiten Zustand.
4. Review des State-of the Art über die Möglichkeiten, bestimmte Charakteristiken für die Positionsbestimmung zu verwenden.

Aus den Meilensteinen ergeben sich die folgenden Forschungsfragen:

- RQI:** Unterscheiden sich die gewählten Metriken (mAP, top-1, pr@10, pr@100) signifikant anhand der verwendeten Binarization und inpainting Methoden?
- RQII:** Unterscheiden sich die gewählten Metriken (mAP, top-1, pr@10, pr@100) signifikant wenn nur der Text oder nur die Fasern als Eingabe verwendet werden im Gegensatz zu den unbearbeiteten Daten?
- RQIII:** Unterscheiden sich die gewählten Metriken (mAP, top-1, pr@10, pr@100) signifikant, wenn zum trainieren zusätzlich künstliche Daten verwendet werden, welche durch ein GAN erzeugt wurden?

Supervisors: Dr.-Ing. V. Christlein, Prof. Dr.-Ing. habil. A. Maier, Mathias Seuret M. Sc.
Student: Timo Bohnstedt
Start: November 8th, 2021
End: April 8th, 2022

References