

# iART

## Eine Suchmaschine zur Unterstützung von bildorientierten Forschungsprozessen

### Schneider, Stefanie

stefanie.schneider@itg.uni-muenchen.de  
Ludwig-Maximilians-Universität München, Germany

### Springstein, Matthias

matthias.springstein@tib.eu  
Technische Informationsbibliothek (TIB), Germany

### Rahnama, Javad

javad.rahnama@uni-paderborn.de  
Universität Paderborn, Germany

### Kohle, Hubertus

hubertus.kohle@lmu.de  
Ludwig-Maximilians-Universität München, Germany

### Ewerth, Ralph

ralph.ewerth@tib.eu  
Technische Informationsbibliothek (TIB), Germany;  
Forschungszentrum L3S, Leibniz Universität Hannover,  
Germany

### Hüllermeier, Eyke

eyke@ifi.lmu.de  
Ludwig-Maximilians-Universität München, Germany

Kunsthistorische Erkenntnisprozesse werden über Ähnlichkeitsbeschreibungen angetrieben: Bei Heinrich Wölfflin finden diese in formanalytischer Perspektive statt, bei dem gleichzeitig tätigen Aby Warburg aus der Sicht des kulturwissenschaftlich forschenden Ikonologen. Beide gelten als Väter der modernen Kunstgeschichte, die deren bis heute aktuellen methodischen Zugriffe im Bereich von Form und Semantik grundgelegt haben. Wölfflin (1915) visualisiert Ähnlichkeiten und Differenzen in seinen „Kunstgeschichtlichen Grundbegriffen“ auf einer Doppelseite mit gegenüberliegenden Vergleichsbeispielen, kategorisiert in fünf binären Gegensätzen. Warburg gestaltet seine zu Berühmtheit avancierten Bilderatlastafeln in Gruppen mit loserer, ähnlichkeitsnaher Verbindung, deren Qualität nicht immer leicht zu erkennen ist (Warnke und Brink 2000; Ohrt et al. 2020). Ein digitales Werkzeug, das ein umfangreiches, nicht über Kanonisierungsprozesse vorselektiertes Bildmaterial nach unterschiedlich gewichteten Ähnlichkeitskriterien zu filtern in der Lage ist, scheint daher in hohem Maße erwünscht – gerade, weil elektronische Bilddatenbanken inzwischen über große Mengen von Reproduktionen verfügen.

Bisherigen Ansätzen fehlt es jedoch entweder an einer Feinabstimmung auf die kunsthistorische Domäne (Rossetto et al. 2016), an flexiblen Suchabfragestrukturen, die sich den Bedürfnissen der Nutzer:innen anpassen (Lang und Ommer 2018), oder an der

Möglichkeit, eigene Datensätze hochzuladen und zu verwalten (Offert, Bell und Harlamov 2021). Mit iART<sup>1</sup> wird der Versuch unternommen, dieses Desiderat mithilfe einer offenen Web-Plattform zu schließen. Das Retrieval von Objekten erfolgt, wie im Weiteren gezeigt wird, nicht nur mit durch Deep Learning generierte Schlagwörter, sondern auch unter Verwendung multimodaler Embeddings, die eine Suche bspw. auf Grundlage detaillierter Szenenbeschreibungen ermöglichen. Eine intuitive Benutzeroberfläche unterstützt die Nutzer:innen bei der Definition von Abfragen und der Untersuchung der Ergebnisse.

## Infrastruktur

Das DFG-geförderte Projekt wurde von 2019 bis 2021 umgesetzt vom Lehrstuhl für Mittlere und Neuere Kunstgeschichte der Ludwig-Maximilians-Universität München, der Forschungsgruppe „Visual Analytics“ der TIB Hannover und der Fachgruppe „Intelligente Systeme und Maschinelles Lernen“ des Heinz Nixdorf Instituts der Universität Paderborn. Es ist geplant, dass die TIB Hannover die entwickelte Plattform auch über die Projektlaufzeit hinaus als Infrastrukturdienst zur Verfügung gestellt und somit dessen Nachhaltigkeit sichert. Weiterhin ist der Quellcode für alle Komponenten frei verfügbar, so dass andere Forscher:innen die Software nachnutzen und erweitern können.<sup>2</sup>

Die beschriebene Software läuft auf zwei Rechnern: Das erste System ist mit einer Grafikkarte (GeForce GTX 1080 Ti) ausgestattet und dient der Indizierung der Daten und der Ähnlichkeitssuche neu hochgeladener Bilder. Weiterhin läuft auf diesem Rechner eine Elasticsearch-Instanz. Das zweite System stellt die Webseite bereit und speichert zu importierende Bilder.

## Backend

Die Aufgabe des Backend besteht sowohl darin, Informationen über Datensätze und Nutzer:innen zu speichern, als auch mithilfe einer API (Application Programming Interface) verschiedene Möglichkeiten des Retrieval zur Verfügung zu stellen. Um die Anpassung an unterschiedliche Forschungsinteressen zu erleichtern, ist die Software modular aufgebaut. Daher sind die einzelnen Indizierungsschritte in Plug-ins ausgelagert und die Benutzerverwaltung von der Suchinfrastruktur getrennt (Abb. 1). Alle Modelle werden mit einem RedisAI-Inferenzserver<sup>3</sup> beschleunigt, um die für die Berechnung benötigten Ressourcen optimal zu verwalten. Dieser Schritt erleichtert es, verschiedene Deep-Learning-Modelle auf einer einzigen Grafikkarte laufen zu lassen und ermöglicht den Einsatz von Backend-Systemen wie PyTorch oder TensorFlow. Die Kommunikation zwischen Indexserver und Frontend wird mithilfe eines auf Python basierenden Django-Webservice umgesetzt.<sup>4</sup> Dieser Service kümmert sich auch um die Verwaltung der Nutzer:innen, von ihnen angelegte Lesezeichen und hochgeladene Bildbestände.

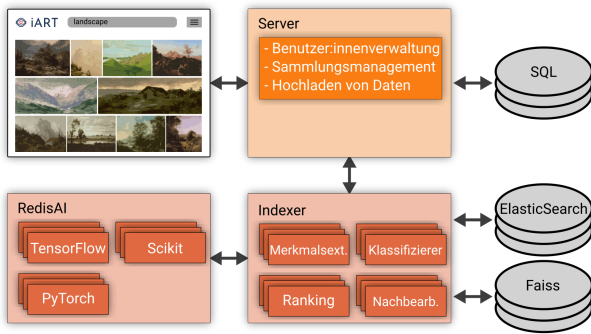


Abb. 1: Architektur mit zugehöriger Datenbankstruktur und RedisAI-Inferenzserver.

## Plug-ins

Zumeist extrahieren Suchmaschinen eine einzige Repräsentation eines Bildes, die für alle Suchanfragen verwendet wird. Kunsthistoriker:innen müssen Objekte jedoch unter verschiedensten Gesichtspunkten finden, etwa in Hinblick auf Komposition oder Farbe. Daher generiert iART eine Vielzahl von Merkmalen pro Bild, die je nach Interesse der Forscher:innen gewichtet werden können. Das zugrundeliegende System wurde flexibel entwickelt, sodass es durch Plug-ins leicht erweitert werden kann: Die einzelnen Plug-ins wurden nach einer Schnittstellendefinition für den jeweiligen Erweiterungstyp in einer eigenen Klasse implementiert, die zur Laufzeit geladen wird. So kann das Framework durch Anlegen einer einzigen Datei und Aktualisieren einer Konfigurationsdatei erweitert werden. Dies gilt insbesondere für die Merkmalsextraktion, die Klassifikation von Bildinhalten, das Ranking der Ergebnisse und verschiedene Nachbearbeitungsschritte, die der Visualisierung und dem Clustering dienen. Die Pipeline ist in Abb. 2 dargestellt. Folgende Plug-ins sind u. a. implementiert:

1. *iMet*. Das Metropolitan Museum of Art stellt im seit 2019 auf Kaggle stattfindenden iMet-Wettbewerb einen Teil seines Bestands für das Training von Bildklassifikatoren zur Verfügung.<sup>5</sup> Dabei sind die 142.119 Bilder des Datensatzes mit 3.474 Tags, etwa aus den Bereichen Region, Kultur und Medium, annotiert. Ein ResNet-Ansatz (He et al. 2016) wird verwendet, um die Schlagwörter vorherzusagen. Dabei erreicht das Modell einen F<sub>2</sub>-Score von 73,6 Prozent auf einem 20-Prozent-Subsample des Datensatzes. Die Metrik wurde für den iMet-Wettbewerb festgelegt und berechnet sich aus der Anzahl der gefundenen Bilder in Relation zu deren richtiger Klassenzuordnung.
2. *Painter by Numbers*. Der ebenso im Rahmen eines Kaggle-Wettbewerbs erstellte Painter-by-Numbers-Datensatz umfasst 80.000 kunstwissenschaftlich relevante Trainingsbilder aus WikiArt, die u. a. mit Genre- und Stilzuschreibungen versehen sind.<sup>6</sup> Das Modell für dieses Plug-in besteht aus einer ResNet-Architektur, die 43 Genre- und 142 Stilvorgaben vorhersagt und dabei eine Genauigkeit von 63,8 bzw. 40,7 Prozent auf einem 20-Prozent-Subsample des Datensatzes erreicht. Ein weiteres Plug-in mit derselben Architektur wird genutzt, um Genre- und Stilmerkmale für die bildgesteuerte Ähnlichkeitssuche zu extrahieren.
3. *BYOL*. BYOL (Bootstrap Your Own Latent; Grill et al. 2020) ist eine selbstüberwachte Lernmethode für neuronale Netze, bei deren Verwendung keine Annotationen für Trai-

ningsbilder benötigt werden. Ziel während des Trainings war es, dass sich zwei Ausschnitte desselben Bildes im Merkmalsraum ähnlicher sein sollten als Ausschnitte anderer Bilder. Die in iART verwendete Version des Modells wurde auf Bildern eines kunsthistorisch spezifischen Wikidata-Samples trainiert.

4. *CLIP*. CLIP (Contrastive Language-Image Pre-training; Radford et al. 2021) beinhaltet Modelle für visuelle und textuelle Informationen, die mithilfe von mehreren Millionen Bild-Text-Paaren trainiert wurden. Ziel des Trainings war es, dass visuelle und textuelle Darstellung für ein Bild-Text-Paar im Merkmalsraum möglichst nah beieinander liegen. Durch die multimodale Ausrichtung wird das Modell in iART auf zwei Weisen eingesetzt: Zum einen werden Suchanfragen mithilfe des Textmodells in den Bildbereich überführt. Dadurch können Nutzer:innen nicht nur nach extrahierten Tags und Metadaten suchen, sondern Bilder auch direkt beschreiben. Zum anderen kann das Modell mit einem Zero-Shot-Ansatz textuelle Schlagwörter vorhersagen, ohne mit diesen trainiert worden zu sein. Wir verwenden diesen Ansatz, um 21.484 Notationen des etablierten Dezimalklassifikationssystems Iconclass (van de Waal 1973–85) zu generieren. Hierzu werden textuelle Merkmale für alle den Notationen zugehörigen Beschreibungen extrahiert und mit den visuellen Embeddings der entsprechenden Bilder verglichen. Im System gespeichert werden alle Labels, die über einem bestimmten Schwellenwert liegen.

Der Einfluss jedes Plug-ins kann mithilfe von Schiebereglern präzise modifiziert werden, um Schwächen einzelner Modelle auszugleichen: Ein generischer ImageNet-Merkmalsextraktor wurde bspw. nur auf wenigen visuellen Konzepten trainiert, sodass eher im Bildhintergrund situierte Phänomene unberücksichtigt bleiben. Konträr dazu berücksichtigt das CLIP-Modell aufgrund der anderen Trainingsmethode wesentlich mehr Informationen des gesamten Bilds.

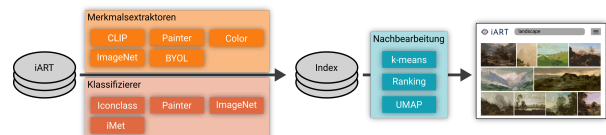


Abb. 2: Indizierungs- und Nachbearbeitungsschritte mit entsprechenden Plug-ins.

## Frontend

Die webbasierte Benutzeroberfläche von iART wurde mit Vue.js<sup>7</sup> erstellt und in JavaScript geschrieben. Sie integriert durch das UI-Framework Vuetify<sup>8</sup> bewusst Googles Material Design. Diese Entscheidung hat zwei Gründe. Zum einen ist Google als Suchstandard etabliert: Plattformen, die sich stark von Google unterscheiden und nicht dessen Usability-Standards entsprechen, sind im Nachteil; wie zuletzt wieder Kröber, Münster und Messmer (2020) gezeigt haben. Zweitens soll der Zugang für Lai:innen nicht unnötig erschwert werden; allein die nicht Metadaten-getriebene Suche kann schließlich als zunächst gewöhnungsbedürftig empfunden werden. Dementsprechend klassisch ist die Positionierung der Einzelkomponenten in iART: Der altbekannte Suchschlitz befindet sich oben in der Mitte, während erweiterte Einstellungsmöglichkeiten in einem Banner darunter erscheinen (Abb. 3, oben).

Verschiedene Objektansichten vereinfachen die Exploration der Suchergebnisse. Standardmäßig wird ein Bildraster angezeigt, über das bei Bedarf weitere Details, wie z. B. Metadaten des jeweiligen Objekts, bereitgestellt werden. Die Ergebnisse können jeweils auf- und absteigend nach Relevanz, Titel oder Entstehungszeitpunkt sortiert werden. Eine Clustering-Visualisierung der Objekte als Bilderkarussells vertikal nach Gruppen getrennt (Abb. 3, Mitte). Der zugrundeliegende Algorithmus ist entsprechend der eigenen Forschungsinteressen konfigurierbar: Unterstützt wird aktuell das partitionierende Verfahren *k-means*; in Zukunft implementiert werden sowohl hierarchische als auch dichte-basierte Ansätze, bspw. *DBSCAN*. Für fortgeschrittene Anwendungsfälle ist es möglich, die Bilder mit der Dimensionsreduktionstechnik *UMAP* (Uniform Manifold Approximation and Projection; McInnes, Healy und Melville 2018) auf einer zweidimensionalen Leinwand anzuordnen, in der farbliche Markierungen die Gruppenzugehörigkeiten der jeweiligen Objekte indizieren (Abb. 3, unten). Durch den in *SciPy* implementierten *Jonker-Volgenant-Algorithmus* können die Bilder zudem überlappungsfrei in einem Raster positioniert werden (Virtanen et al. 2020, Crouse 2016). Zoom- und Filteroperationen, etwa ein interaktives *Drag-Select* zur Gegenüberstellung mehrerer Objekte, unterstützt *iART* mithilfe der Bibliothek *vis.js*.<sup>9</sup>

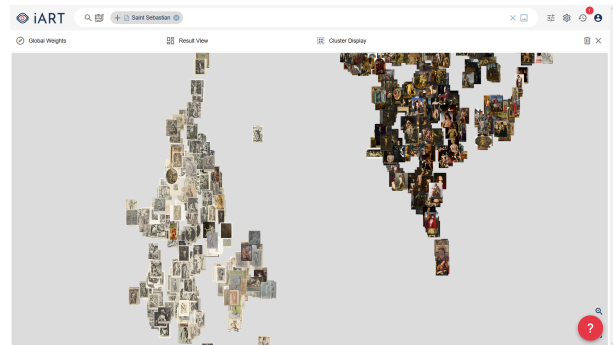
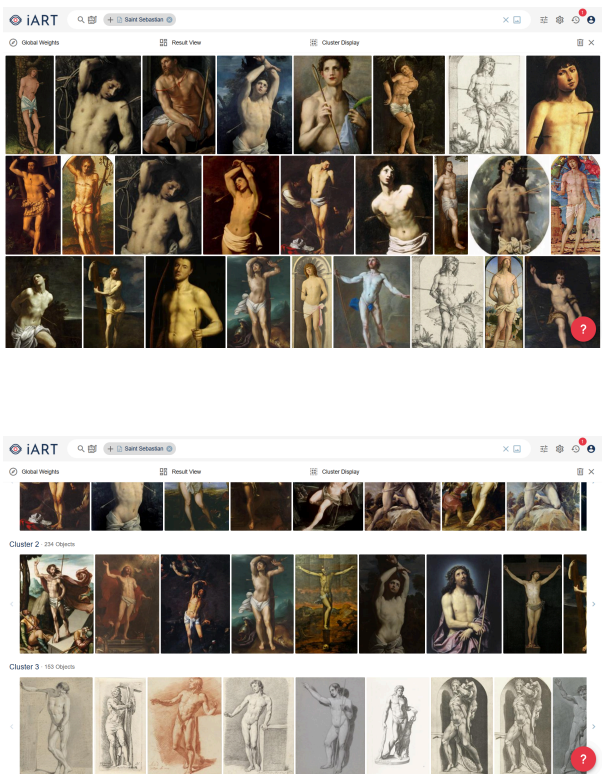


Abb. 3: Verschiedene Objektansichten zur Exploration der Suchergebnisse. Standardbildraster (oben), nach Gruppen getrennte Bilderkarussells (Mitte), Anordnung auf einer zweidimensionalen Leinwand (unten).

## Datensätze

*iART* integriert ein breites Spektrum offen lizenzierter Bildinventare, das fortlaufend erweitert wird. Momentan bereitgestellt werden Daten aus fünf kunsthistorisch relevanten Quellen: des niederländischen Rijksmuseums (392.624 Objekte), der Wikidata (347.448 Objekte), des virtuellen Münzkabinetts *KENOM* (119.580 Objekte), der Social-Tagging-Plattform *AR-Tigo* (54.411 Objekte; Becker et al. 2018) und des Museumportals *Kulturerbe Niedersachsen* (12.085 Objekte).<sup>10</sup> Die Objekte wurden entweder mittels *Web Scraping* oder über offiziell verfügbare APIs extrahiert. Demnächst folgen u. a. Bestände des *Musée du Louvre* und *Victoria and Albert Museums*.<sup>11</sup> Zum Metadaten-gestützten Retrieval offeriert *iART* im Frontend eine facettierte Suche, die die Objekte nach von den jeweiligen Institutionen vorgehaltenen Kategorien, z. B. Genre oder Medium, unscharf filtert (Abb. 4). Die Kategorien wurden manuell auf ein gemeinsames Schema überführt.

Um zu gewährleisten, dass selbst spezifische kunsthistorische Forschungsanliegen flexibel adressiert werden können, ist der Import von eigenen Datenbeständen für registrierte Nutzer:innen möglich. Zum einen werden bspw. als *CSV*-Datei bereitgestellte Metadaten für die facettierte Suche nutzbar gemacht, zum anderen in einer *ZIP*-Datei gebündelte Bildinhalte mit den zuvor beschriebenen *Plug-ins* analysiert. Anschließend können Nutzer:innen ihre hochgeladenen Sammlungen einzeln oder im Kontext mit frei verfügbaren Inhalten untersuchen. Damit ist eine *Deep-Learning*-gestützte Suche auch für Lai:innen praktikabel, die nicht nur auf übliche Schnittstellen, wie *Googles Cloud Vision API*,<sup>12</sup> zurückgreifen möchten.

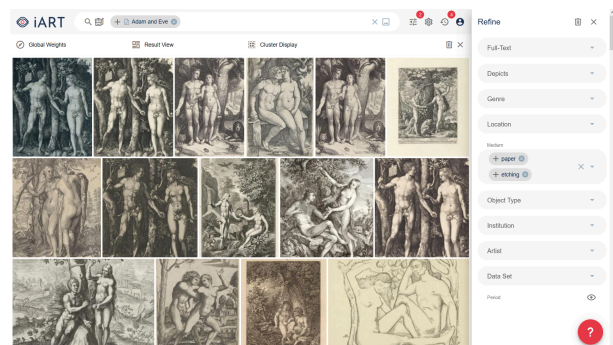




Abb. 4: Beispiel für eine nach Medium facettierte Suche mit der Abfrage „Adam and Eve“.

## Use Cases

Die Vorteile von iART erschließen sich insbesondere bei Suchanfragen, die aufgrund ihrer semantischen Komplexität bislang nahezu unmöglich waren – oder nur in äußerst feingranular verschlagworteten Systemen sinnvolle Ergebnisse bringen. Die textbasierte Suche „Death of Marat“ gibt bspw. vier relevante Ergebnisgruppen zurück: erstens Jacques-Louis Davids „Der Tod des Marat“ (1793) in verschiedenen Reproduktionen, dazu weitere Beispiele für diese Ikonographie; zweitens Beispiele für Darstellungen des toten Christus, auf den sich auch die Marat-Darstellungen beziehen; drittens andere Figuren, denen der Arm in ähnlicher Weise herabhängt wie dem David’schen „Marat“; und viertens völlig andere Ikonographien, die formal – etwa in der Anordnung der gebogenen Linie vom Oberkörper des toten Marats und seinem Arm – auf den David’schen „Marat“ bezogen werden könnten (Abb. 5, oben).<sup>13</sup> Wir sehen vor allem drei Anwendungsszenarien, die im Folgenden exemplarisch beschrieben werden:

1. *Suche nach ähnlichen Ikonographien.* Die Abfrage mit einem Bild von Leonardo da Vincis „Salvator Mundi“ (um 1500) führt zu einer Reihe von Salvator-Darstellungen, unter die sich nur wenige andere Ikonographien mischen.<sup>14</sup> Dabei werden vor allem auch Objekte gefunden, die anders – z. B. in einer anderen Sprache – bezeichnet werden. Die vom System ermöglichte Clusterung mit k-means trägt zu einer weiteren Präzisierung der Ergebnismenge bei.
2. *Suche zur Rekontextualisierung.* Eine Recherche auf Basis von Vincent van Goghs „Die Kartoffelesser“ (1885) als Referenzbild weist auf spezifisch holländische bzw. nordeuropäische Darstellungen bäuerlicher Mahlzeiten und liefert eine empirische Begründung für bekannte Vermutungen zur Geschichte der Genreverteilung in der europäischen Malerei.<sup>15</sup>
3. *Suche im historischen Umkreis.* Eine Query-by-Image-Suche mit Tizians „Jacopo de Strada“ (1567/68) resultiert in einer Reihe von Halbfigurenporträts, die häufig aus dem venezianisch-norditalienischen Raum stammen (Abb. 5, unten).

<sup>16</sup> Warum das so ist, lässt sich kaum eindeutig bestimmen: Es könnte mit Gestaltungseigenheiten zusammenhängen, dürfte letztendlich aber ein Beispiel für die Blackbox des Deep Learning sein. Gerade hier erweist sich die facettierte Suche mit ihrer Möglichkeit, die Objekte nach Medium einzuschränken, als sinnvoller Filtermechanismus, bei dem die Konzentration im norditalienischen Bereich am deutlichsten zur Geltung kommt.

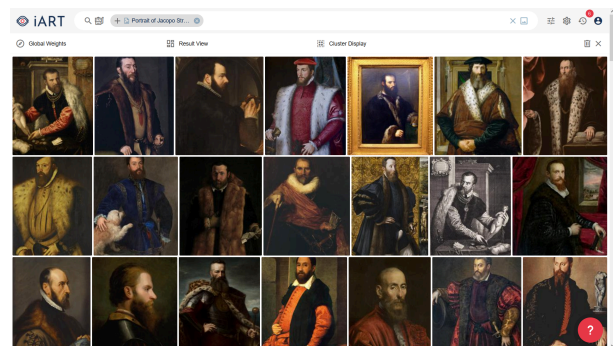
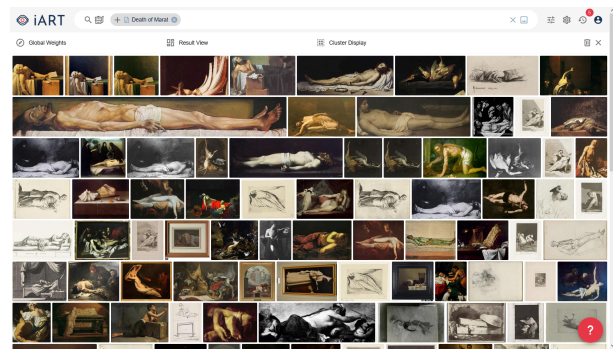


Abb. 5: Ergebnisse für eine Query-by-Text-Suche nach „Death of Marat“ (oben) und eine Query-by-Image-Suche nach Tizians „Jacopo de Strada“ (unten).

## Fazit

Die Anwendungsszenarien zeigen, dass iART als unterstützendes Werkzeug für die kunst- und kulturwissenschaftliche Forschung dienen kann, indem es für eine Forschungsfrage interessante Bildobjekte identifiziert, extrahiert und analysiert. Da das System verschiedene Klassifizierungs-Plug-ins und Feature-Extraktoren unterstützt, können Nutzer:innen es an ihre Bedürfnisse anpassen. Auch die maschinell gesteuerte Suche sollte dabei prinzipiell als Anreiz zur weiteren Exploration verstanden werden und nicht als Instrument, das per se perfekte Ergebnisse liefert. Gerade durch zunächst „unsinnig“ oder offensichtlich „falsch“ erscheinende Resultate können sich durchaus näher zu begutachtende Forschungsperspektiven ergeben.

## Danksagung

Das Projekt „iART: Ein interaktives Analyse- und Retrieval-Tool zur Unterstützung von bildorientierten Forschungsprozessen“ wurde von der Deutschen Forschungsgemeinschaft (DFG) gefördert (Projektnummer 415796915).

## Fußnoten

1. <https://iart.vision>, letzter Zugriff 30. November 2021.
2. <https://github.com/TIBHannover/iart>, letzter Zugriff 30. November 2021.



3. <https://oss.redis.com/redisai/>, letzter Zugriff 30. November 2021.
4. <https://www.djangoproject.com/>, letzter Zugriff 30. November 2021.
5. <https://www.kaggle.com/c/imet-2021-fgvc8>, letzter Zugriff 30. November 2021.
6. <https://www.wikiart.org/de> und <https://www.kaggle.com/c/painter-by-numbers/data>, jeweils letzter Zugriff 30. November 2021.
7. <https://vuejs.org/>, letzter Zugriff 30. November 2021.
8. <https://vuetifyjs.com/en/>, letzter Zugriff 30. November 2021.
9. <https://visjs.org/>, letzter Zugriff 30. November 2021.
10. <https://www.rijksmuseum.nl/en/>, <https://www.wikidata.org/>, <https://www.kenom.de/>, <https://www.artigo.org/> und <https://kulturerbe.niedersachsen.de/>, jeweils letzter Zugriff 30. November 2021.
11. <https://www.louvre.fr/en> und <https://www.vam.ac.uk/>, jeweils letzter Zugriff 30. November 2021.
12. <https://cloud.google.com/vision>, letzter Zugriff 30. November 2021.
13. <https://iart.vision/search?query=%2Btxt%3ADeath+of+Marat>, letzter Zugriff 30. November 2021.
14. <https://iart.vision/search?query=%2Bidx%3A3fa6b53c7e163ebd9663a01ab3efd24a>, letzter Zugriff 30. November 2021.
15. <https://iart.vision/search?query=%2Bidx%3A2e0d7a-d7a1733fdbbad8134b871dd8b0>, letzter Zugriff 30. November 2021.
16. <https://iart.vision/search?query=%2Bidx%3A673681d267163d84900c41c77ce951e2>, letzter Zugriff 30. November 2021.

## Bibliographie

- Becker, Matthias / Bogner, Martin / Bross, Fabian / Bry, François / Campanella, Caterina / Commare, Laura / Cramerotti, Silvia / Jakob, Katharina / Josko, Martin / Kneißl, Fabian / Kohle, Hubertus / Krefeld, Thomas / Levushkina, Elena / Lücke, Stephan / Puglisi, Alessandra / Regner, Anke / Riepl, Christian / Schefels, Clemens / Schemainda, Corina / Schmidt, Eva / Schneider, Stefanie / Schön, Gerhard / Schulz, Klaus / Siglmüller, Franz / Steinmayr, Bartholomäus / Störkle, Florian / Teske, Iris / Wieser, Christoph (2018): *ARTigo. Social Image Tagging [Dataset and Images]* 10.5282/ubm/data.136 .
- Crouse, David F. (2016): „On Implementing 2D Rectangular Assignment Algorithms“, in: *IEEE Transactions on Aerospace and Electronic Systems* 52(4) 10.1109/TAES.2016.140952 .
- Grill, Jean-Bastien / Strub, Florian / Altché, Florent / Tallec, Coentín / Richemond, Pierre H. / Buchatskaya, Elena / Dörsch, Carl / Pires, Bernardo Ávila / Guo, Zhaohan Daniel / Azar, Mohammad Gheshlaghi / Piot, Bilal / Kavukcuoglu, Koray / Munos, Rémi / Valko, Michal (2020): „Bootstrap Your Own Latent. A New Approach to Self-Supervised Learning“, in: *Proceedings of Advances in Neural Information Processing Systems* 21271–21284.
- Ohrt, Roberto / Heil, Axel / The Warburg Institute / Haus der Kulturen der Welt (Eds., 2020): *Aby Warburg. Bilderatlas Mnemosyne. The Original*. Berlin: Hatje Cantz.
- He, Kaiming / Zhang, Xiangyu / Ren, Shaoqing / Sun, Jian (2016): „Deep Residual Learning for Image Recognition“, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778.
- Kröber, Cindy / Münster, Sander / Messemer, Heike (2020): „Bildrepositorien und Forschung mit digitalen Bildern im Bereich der Kunstgeschichte“, in: Schöch, Christof (ed.): *DHd 2020. Spielräume. Digital Humanities zwischen Modellierung und Interpretation. 7. Jahrestagung des Verbands Digital Humanities im deutschsprachigen Raum* 87–90 10.5281/zenodo.3666690 .
- Lang, Sabine / Ommer, Björn (2018): „Attesting Similarity. Supporting the Organization and Study of Art Image Collections with Computer Vision“, in: *Digital Scholarship in the Humanities* 845–856.
- McInnes, Leland / Healy, John / Melville, James (2018): *UMAP. Uniform Manifold Approximation and Projection for Dimension Reduction* <https://arxiv.org/pdf/1802.03426.pdf> [letzter Zugriff 14. Juli 2021].
- Offert, Fabian / Bell, Peter / Harlamov, Oleg (2020): *imgs.ai* <https://imgs.ai/> [letzter Zugriff 14. Juli 2021].
- Radford, Alec / Wook Kim, Jong / Hallacy, Chris / Ramesh, Aditya / Goh, Gabriel / Agarwal, Sandhini / Sastry, Girish / Askell, Amanda / Mishkin, Pamela / Clark, Jack / Krueger, Gretchen / Sutskever, Ilya (2021): *Learning Transferable Visual Models From Natural Language Supervision* <https://arxiv.org/pdf/2103.00020.pdf> [letzter Zugriff 14. Juli 2021].
- Rossetto, Luca / Giangreco, Ivan / Tanase, Claudiu / Schuldt, Heiko (2016): „vitivr. A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections“, in: *Proceedings of the 24th ACM International Conference on Multimedia* 1183–1186.
- van de Waal, Henri (1973–85): *Iconclass. An Iconographic Classification System. Completed and Edited by L. D. Couprie with R. H. Fuchs*. Amsterdam / Oxford / New York: North-Holland Publishing Company.
- Virtanen, Pauli / Gommers, Ralf / Oliphant, Travis E. / Haberland, Matt / Reddy, Tyler / Cournapeau, David / Burovski, Evgeni / Peterson, Pearu / Weckesser, Warren / Bright, Jonathan / van der Walt, Stéfan J. / Brett, Matthew / Wilson, Joshua / Millman, K. Jarrod / Mayorov, Nikolay / Nelson, Andrew R. J. / Jones, Eric / Kern, Robert / Larson, Eric / Carey, CJ / Polat, İlhan / Feng, Yu / Moore, Eric W. / VanderPlas, Jake / Laxalde, Denis / Perktold, Josef / Cimrman, Robert / Henriksen, Ian / Quintero, E.A. / Harris, Charles R. / Archibald, Anne M. / Ribeiro, António H. / Pedregosa, Fabian / van Mulbregt, Paul / SciPy 1.0 Contributors (2020): „SciPy 1.0. Fundamental Algorithms for Scientific Computing in Python“, in: *Nature Methods* 17(3): 261–272.
- Warnke, Martin / Brink, Claudia (Eds., 2000): *Aby Warburg. Gesammelte Schriften. Der Bilderatlas Mnemosyne*. II(1), Berlin: Akademie Verlag.
- Wölfflin, Heinrich (1915): *Kunstgeschichtliche Grundbegriffe. Das Problem der Stilentwicklung in der neueren Kunst*. München: Bruckmann.