Historical Printed Ornaments: Dataset and Tasks

Sayan Kumar Chaki*[1][0000−0003−1390−1329], Zeynep Sonat Baltaci*[2][0000−0001−6749−2407], Elliot Vincent[2,3][0009−0001−1713−2590], Remi Emonet[1,4][0000−0002−1870−1329], Fabienne Vial-Bonacci[5][0000−0001−6202−6407], Christelle Bahier-Porte[5][0000−0003−1775−0142], Mathieu Aubry[2][0000−0002−3804−0193], and Thierry Fournel[1][0000−0002−1613−4594]

[1] Univ. Lyon, UJM-St-Etienne, CNRS, Institut d'Optique Graduate School, Inria, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France

[2] LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

[3] Inria, Ecole normale supérieure, CNRS, PSL Research University, Paris, France

[4] IUF University Institute of France

[5] UJM-St-Etienne, CNRS, IHRIM 5317, Saint-Etienne, France

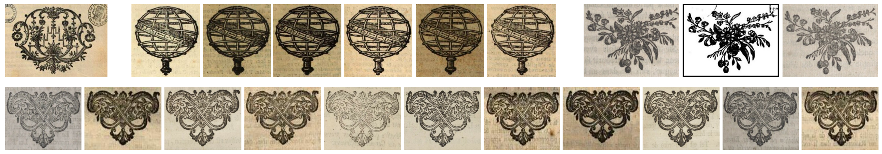
{sayan.kumar.chaki,remi.emonet,fabienne.vial,christelle.porte, fournel}@univ-st-etienne.fr, {sonat.baltaci, elliot.vincent,mathieu.aubry}@enpc.fr

Abstract. This paper aims to develop the study of historical printed ornaments with modern unsupervised computer vision. We highlight three complex tasks that are of critical interest to book historians: clustering, element discovery, and unsupervised change localization. For each of these tasks, we introduce an evaluation benchmark, and we adapt and evaluate state-of-the-art models. Our *Rey's Ornaments dataset* is designed to be a representative example of a set of ornaments historians would be interested in. It focuses on an XVIIIth century bookseller, Marc Michel Rey, providing a consistent set of ornaments with a wide diversity and representative challenges. Our results highlight the limitations of state-of-the-art models when faced with real data and show simple baselines such as k-means or congealing can outperform more sophisticated approaches. Our dataset and code can be found at https://github.com/adress-here. TODO: ask Thierry if he wants a specific link, otherwise Sonat clone and share it on your github

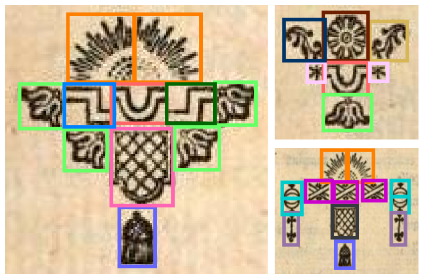
Keywords: Book ornaments · Clustering · Element discovery · Unsupervised change localization

1 Introduction

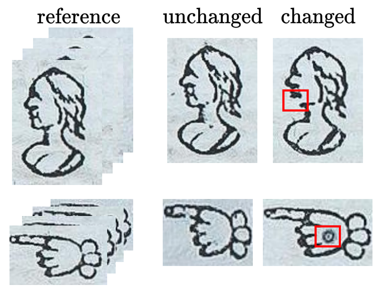
Typographical ornamentation is a key component of historical printed texts. While ornaments were first collected to improve the attribution of books to a particular printer [20, 85, 59], they turned out to be a critical part of material bibliography and book archaeology [68]. Because of the massive amount of available material and the difficulty of annotations, many aspects of the study of book ornaments could benefit from modern unsupervised computer vision. In



(a) Clustering



(b) Element discovery



(c) Unsupervised change localization

Fig. 1: **Our Rey's Ornaments dataset.** Our dataset, based on ornaments found in the books published by or attributed to Marc Michel Rey (1720-1780), focuses on three unsupervised computer vision tasks that are of interest to book historians: (a) image clustering of ornaments printed using woodblocks, (b) unsupervised discovery of elements composing ornaments printed using multiple types, and (c) unsupervised change localization in vignette series.

this paper, we identify three tasks of particular interest - clustering, element discovery, and unsupervised change localization, illustrated in Fig. 1 - for which we built datasets and evaluated state-of-the-art methods, showing that despite the apparent simplicity of the 2D patterns to analyze, these tasks remain extremely challenging and would benefit from more attention.

The challenges come from three main reasons. First, book ornaments are complex objects. Older ornaments come from unique woodblocks, that might have similar appearances but that historians want to differentiate. More recent ornaments are assembled from several vignettes produced by typographical metal types. Such composite ornaments may include tens of vignettes, which one would like to identify as different visual elements, and possibly relate to catalogs that were used to sell metal types. Second, the appearance of each ornament has a lot of variations. It can be due to many factors, including of course degradation of the books and image acquisition conditions, but also aging or degradation of the blocks, 3D effects and variations in the vignettes assembling, and randomness in inking, in the transfer of the ink to different papers and the hand press inking process. Third, the tasks that are of actual interest to book historians are challenging in themselves: no supervision, very few samples corresponding to each ornament or vignette, very imbalanced data, and the importance of fine differences.

Because they are of interest to book historians, many repositories of ornaments already exist, including some online databases [30, 18, 58, 19, 17, 84, 9, 6]. The need for automatic retrieval tools quickly emerged [8]. Some have also been designed for related data, such as ornamental letters [4] or woodcut illustrations [16]. The use of the Visual Image Search Engine [23], which leverages bags of visual words [78, 2], is particularly widespread. However, more advanced tools than image search are scarce. We believe providing relevant annotated data is key for their development.

We built our dataset based on the books published by Marc Michel Rey (1720-1780), the leading publisher of Enlightenment philosophers, who is known to be especially attentive to the quality of his books [67, 3]. The XVIIIth century is a particularly interesting period from the book-ornament point of view since it marks the transition between the dominant use of woodcuts and typographic metal types. Restricting ourselves to a single bookseller is motivated both by historical considerations, the history of this particular bookseller being of interest, and practical ones since this leads to a limited vocabulary of vignettes and woodblocks which makes annotations possible with the help of historian experts.

Our Rey's Ornaments dataset is composed of three parts, composed of distinct image sets, giving insights into our three tasks:

- Our clustering dataset, based on woodblock ornaments, includes 167 images of 36 different ornaments, each associated with 3 to 14 occurrences. We found that DTI clustering [61] outperformed state-of-the-art clustering approaches by a large margin on a balanced subset of images, but that the k-means algorithm [57] on the foundation features (e.g., CLIP [66]) was on-par with DTI-clustering on the full imbalanced dataset. However, all algorithms led to less than 80% accuracy in this setting.
- Our element discovery dataset includes 100 images of composite ornaments containing 1271 elements, from a dictionary of 72 different vignettes. We found all unsupervised element discovery methods to perform poorly. We believe this stresses the potential benefit of our dataset for the community, compared to the synthetic datasets often used to evaluate unsupervised object segmentation methods.
- Our unsupervised change localization dataset includes 30 types of vignettes with four reference instances and two test instances, a normal one and one where changes have been annotated. We found that direct reconstruction-based approaches, such as congealing or VAE-based, performed poorly compared to human annotations, mostly because they are confused by the variations in inking. We see this as an invitation to better formalize the notion of changes relevant to book historians and design-associated algorithms.

The paper is organized as follows: Section 2 reviews related work for our three tasks, Section 3 presents our Rey's Ornaments dataset, tasks, and metrics, and Section 4 discusses the performance of state-of-the-art algorithms for each task. Both our dataset and code will be released upon publication.

2 Related work

Image clustering. The classical k-means algorithm [57] remains the basis for many recent clustering methods. It splits a collection of images into k clusters by jointly optimizing k centroids and the assignment of each image to the closest centroid. Distances between images can be directly computed in pixel space, for example in Transformation Invariant methods [33, 32, 61]. More commonly, k-means is used with learned features, either optimized for a pretext task or together with the k-means clustering. For example, DCN [87] optimizes an autoencoder both for reconstruction and clustering in latent space, CCNN [40] fine-tunes a pre-trained network to minimize a mini-batch k-means loss, and DeepCluster [12] learns features in a self-supervised way using k-means clusters as class labels.

There are of course many different clustering methods. Recently, many approaches defined a loss that is simply minimized with stochastic gradient descent in a deep learning framework. In some, clustering and feature learning are optimized jointly, for example with a loss based on KL divergence [86, 38], mutual information [41, 44], consensus [42, 60], or image likelihood [51]. This can also be done with two-step approaches, first learning features and then optimizing a clustering objective on these features, for example in SCAN [81].

Although clustering methods are adopted in various fields in cultural heritage [29], especially in document analysis, the focus mainly remained on texts rather than visual clues [39, 34]. We evaluate different types of approaches for our dataset: k-means on pixels, transformation invariant k-means [61], k-means on pre-trained features, joint feature and cluster learning with mutual information [44] and consensus objectives [60], as well as optimizing a clustering objective on self-supervised features [81].

Element discovery. By element discovery, we mean identifying different categories of visual elements and decomposing images into such elements without supervision. This problem is related to object co-segmentation and discovery, which has been addressed for example by using visual words and a topic hierarchy [79, 72, 11] or by computing similarities between image regions [36, 47, 70, 15, 83, 76]. However such methods are designed to work with textured and discriminative regions and are unlikely to work with our composite ornaments.

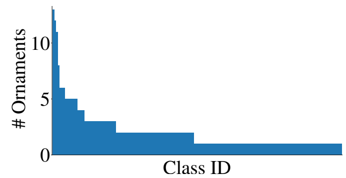
Element discovery is also related to what has recently been referred to in deep learning as unsupervised multi-object segmentation. However, these methods do not always model background and often do not model any notion of class or category for the discovered segments. We give an overview of these models following the classification introduced in [49]. *Pixel-Space Approaches*, such as [10, 25, 24], model images using a predefined number of objects, and determine per-pixel allocations to objects without discriminating two occurrences of the same object. An additional limitation is that these approaches are typically computationally intensive and restricted to a limited number of objects. Thus we do not evaluate them, but some of their core principles are integrated into the methods presented below. *Glimpse-Based Approaches* [27, 22, 55, 45, 90, 73] follow the seminal AIR [27] method and perform element prediction based on regions from the input image, referred to as "glimpses". We evaluated the SPACE [55] and AST-argmax [73] approaches, which have the advantage of having a background model. However they do not differentiate classes of elements, thus we combine them with feature clustering to obtain categories on the extracted elements. To the best of our knowledge, the only method in this category that identifies classes is GMAIR [90]. We thus also evaluate it, together with the SPAIR [22] method it builds on, but both methods suffer from not having a background model. *Sprite-based approaches* [80, 62] decompose images into elements by learning representative prototypes and their transformation to optimize a reconstruction loss. We evaluate the DTI-sprite approach [62], because it models elements scales, which is critical for our data.

These unsupervised multi-object segmentation methods are mostly evaluated quantitatively only on synthetic datasets such as Tetrominoes [37], CLEVR6 [46, 37], multi-dSprites [48], or ClevrTex [49], where performances are typically very good. However, they have been shown to perform very poorly on natural 3D world images [88]. Because real applications are unclear, the evaluations often focus mainly on instance segmentation, i.e., not considering semantics. Thus, we believe the introduction of a real dataset, challenging but simpler than natural scenes, with a clear associated task, can be a significant contribution to this entire field and improve the benchmarking of future methods.

Unsupervised change localization. By unsupervised change localization, we refer to methods that identify pixels where changes occur in a test image compared to a reference, which can also be referred to as anomaly or novelty detection in the literature. The reference can be a single image sample or a small collection of samples. Many methods address this task, and a complete review is out of the scope of this work, we refer the reader to existing reviews and benchmarks [65, 7, 64, 71]. Approaches can be broadly separated between reconstruction-based and feature-based methods. To localize anomalies or changes, the reconstruction-based methods compute a pixel-wise difference between the reconstructed image and input. They usually need class-dependent thresholds and anomalous training images. This motivated the development of feature-based methods such as [82]. Such methods compute a false-color heatmap from the extracted features to highlight regions containing (part of) anomalies [56]. Feature-based methods typically do not precisely localize the changed pixels, we thus focus on reconstruction-based methods.

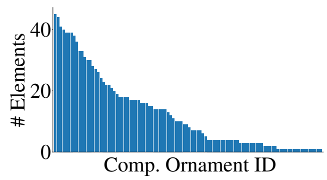
The reconstruction can be either directly computed from the reference samples or produced by a deep neural network. We consider as baseline simply comparing the test image to the average of the reference images, or solving for the joint alignment problem before computing the average, which can be seen as a variation of the classical congealing problem [53, 21]. Deep reconstruction approaches are typically based either on GANs [74, 89, 1] or autoencoders [50, 5], which in theory can capture complex variability in the reference examples. We benchmark both a classical VAE-based method [50] and a more

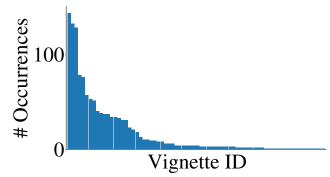
Clustering	Base	Imbalanced	Balanced
# ornaments	339	167	70
# classes	163	36	14



(a) Clustering dataset

Element Discovery	
# comp. ornaments	100
# vignettes	72
# vignettes occ.	1271





(b) Element discovery dataset

Table 1: **Dataset statistics.** Our clustering and element discovery datasets are highly imbalanced, which is one of the key challenges of real data but rarely considered in benchmarks.

recent autoencoder-based method that leverages spatial transformers [43] to obtain sharper reconstructions [14].

Change localization datasets exist in various domains such as 3D-MR-MS [54] in medical imaging, S2Looking [75] in remote sensing, CDnet [35] in video surveillance, MVTEC [7] in industrial inspection, and TAMPAR [63] in tampering detection. To the best of our knowledge, no such dataset exists for historical ornaments, while change localization is one of the key steps in analyzing printed material, and inking variability makes this task very specific.

3 The Rey's ornaments dataset

This section presents our datasets targeted toward clustering, element discovery, and unsupervised change localization. The ornaments for each task are extracted from books listed in the Marc-Michel Rey database [67] and shared by the Bibliothèque Nationale de France (BnF) or the Bibliothèque Municipale de Lyon (BML). All the annotations were made under the supervision of book historians.

3.1 Block ornaments and clustering

The clustering dataset contains ornaments printed from woodcut blocks. Examples are shown in Figure 1a. We considered an initial set of 339 images of block ornaments, we refer as the base set, and annotated their class label with the help of the VISE [23] similarity search engine. This led to 163 ornament classes, most

Fig. 2: Examples of composite ornaments from our synthetic dataset.

of them corresponding to one or two images. While we will release this full set and the associated annotations, we focused our evaluation on two subsets:

- an imbalanced subset, with all the 167 images from the 36 classes that have at least 3 instances,
- a balanced subset, with 70 images, built by randomly sampling 5 images from the 14 classes that have at least 5 instances.

This choice was motivated by the fact that most existing clustering approaches do not handle well large number of classes with a single or very few samples. The statistics of our dataset are visualized in Figure 1a.

Evaluation metrics. We use two standard metrics to evaluate our clustering: normalized mutual information (NMI) and accuracy. Following the standard practice to evaluate clustering, we compute accuracy by matching clusters and classes using the Hungarian matching algorithm [52].

3.2 Composite ornaments and element discovery

The element discovery dataset contains 100 images of composite ornaments. In this set, we first identified 72 categories of vignettes as composition elements. We identified 51 of those in two vignette catalogs by providers of metal types in use in Rey's publishing [26, 69]. For each each category of vignette we selected a representative example that we did not use in our experiments but was the reference for annotation and that we will release as part of the dataset. We then manually annotated in each composite ornament the bounding box and class of each element. Three examples of composite ornaments annotated with semantic bounding boxes for each vignette can be seen in Figure 1b. The statistics of the resulting dataset are presented in Table 1b. Our 100 composite ornaments contained a total of 1271 elements each corresponding to one of the 72 vignettes. Some ornaments included a single element, others more than 40. Some vignettes were used a single time, others more than 100 times. This strong imbalance is part of the challenge of our dataset and representative of statistics encountered in real problems. Another challenge is that the elements are often grouped, very close, or even touching each other, and are thus much more challenging to separate than the objects in the synthetic unsupervised object segmentation datasets [37, 46, 48, 49].

Synthetic dataset. Because we found that the existing algorithms trained directly on our dataset performed very poorly, we also created a synthetic dataset, more similar to the ones existing in the literature. We first selected 67 distinct parchment images as empty backgrounds, then pasted on these backgrounds up to 10 elements uniformly sampled from our dictionary at random locations. Examples of the resulting synthetic composites are shown in Figure 2. We used these simpler examples, without annotations, to pre-train algorithms, and then fine-tuned them on the real data, which resulted in improved performances (see Section 4.2).

Evaluation metrics. Recent unsupervised object segmentation methods typically focus their evaluations on *instance* segmentation, often discarding background pixels. However, we want to identify the classes of the elements as well as their localization. Coarsely localizing them is also sufficient for applications, while annotating exact segmentation would be costly. We thus turn to an object detection metric, namely mean average precision (mAP). Following standard practices [28], we consider a detection to be accurate if its Intersection over Union (IoU) with the annotation is larger than 0.5. To match discovered elements categories we follow the same approach as for clustering algorithms evaluation and use the Hungarian matching algorithm [52] to match the discovered categories with vignettes categories, similar to [77]. To analyze separately the influence of the elements' localization and identification, we also measure class-independent element detection and report the results using average precision (AP).

3.3 Vignettes' series and unsupervised change localization

The change localization dataset contains 180 vignettes' images from the catalogs published by Enschede [26] and Rosart [69], two providers of metal types used by Rey to build composite ornaments, as well as in one published by Fournier [31] which was used by Rey's counterfeiters. In such catalogs, the same vignette category is printed on one or multiple lines containing several occurrences. This enabled the publishers to see variations that could be expected for a given vignette but also enabled us to easily find variations, i.e., vignette instances that we perceive as different from the other ones in the same category. Inspecting these catalogs, we built a dataset containing 30 different classes of vignettes, each one associated with 4 images of normal print we used as reference, and 2 test images, one associated with a normal print and one associated with a print error, which we refer to as 'unchanged' and 'changed' examples, respectively. Examples of this dataset are shown in Figure 1c. In the ' changed' image, we annotated the binary mask of the pixelwise regions corresponding to the perceived difference.

Evaluation metrics. We evaluate change localization by computing Intersection ever Union (IoU) between annotated and predicted changes on each image. We compute this metric either only on the changed image or on both the changed

Dataset		Imbalanced Dataset		Balanced Dataset	
Method		Acc. (%) ↑	NMI (%) ↑	Acc. (%) ↑	NMI (%) ↑
clustering over features					
IIC [44]		19.2	40.6	25.0	35.0
SCAN [81]		46.1	68.7	47.1	64.6
DivClust [60]		54.1±0.5	80.5±0.3	67.0±2.1	78.6±0.7
	Feat. Extractor				
k-means [57]	DINO-ViTB16[13]	73.9±0.9	89.9±0.4	70.9±3.3	86.8±1.6
k-means [57]	CLIP-RN50x16[66]	74.6±3.0	90.1±1.1	78.3±2.9	89.8±1.2
clustering over pixels					
k-means [57]		65.6±1.8	84.1±1.1	74.3±1.9	83.9±0.7
DTI Clustering [61]		**75.7±0.8**	**90.7±0.6**	**87.4±1.9**	**92.1±1.2**

Table 2: **Image clustering baselines.** We report the clustering accuracy (Acc.) and normalized mutual information score (NMI) on our imbalanced and balanced datasets. We report the standard error among 5 runs for the fastest methods, outline the best results in bold, and underline ones within one standard error.

and unchanged images - which is harder since the method has to correctly recognize that none of the variations in the unchanged images are significant. We then report a mean Intersection over Union (mIoU) over the 30 vignette classes.

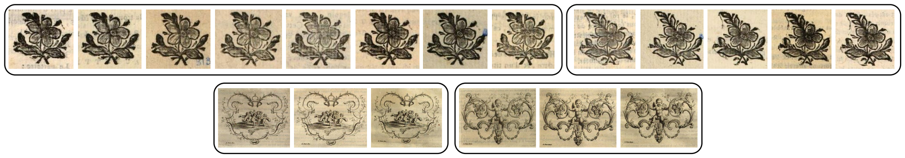
4 Results and analysis

In this section, we analyze the results of diverse methods for each of our tasks. This provides insight both into the specific challenges of our problems and the limitations and strengths of state-of-the-art algorithms.

4.1 Block ornaments and clustering

Methods We tested both methods that performed clustering on pixels and features. Using the raw pixel values and the standard L2 distance, we evaluated k-means [57], and Deep Transformation-Invariant Clustering [61] which jointly learns and transforms cluster centers to reconstruct images. For clustering on features, we used both pre-trained standard features and methods that specifically learn features for clustering on a specific dataset. As examples of standard pre-trained features, we used the self-supervised DINO [13] features and the supervised CLIP [66] features. As methods that jointly train for features and clustering, we evaluated ICC [44], SCAN [81], and DivClust [60].

Results Our quantitative results are reported in Table 2 and highlight several facts. First and most surprisingly, simply performing k-means on pixel values



(a) Valid clusters obtained with DTI clustering.

(b) Failure examples obtained with DTI clustering.

Fig. 3: **Qualitative results for clustering.** Although dlusters obtained DTI clustering are often valid (a), there are failure cases (b) due for example to similar vignettes (top), or split clusters (bottom). Results are qualitatively similar when using k-means with pretrained feature extractors.

performed better than method learning ad-hoc features and clustering on both the balanced and imbalanced datasets. We believe that can be explained by the fact that our images are relatively aligned and similar in appearance (dark ink on light paper), and by the fact that our datasets have very few examples per class. This highlights a clear limitation of common benchmarks and the complex methods that report state-of-the-art performances on them. Second, performing k-means over pre-trained features performs better than any other feature-based method, in particular for the imbalanced dataset, with a small advantage for the CLIP features compared to the DINO features. Third, DTI clustering consistently improves over all approaches, but the margin is small on the imbalanced dataset.

Figure 3 shows some qualitative results obtained with the best-performing method, DTI clustering, on the imbalanced dataset. Valid clusters are obtained even with variations in inking and paper appearance (a). Failure cases are typically related to similar-looking but different ornaments being grouped in the same cluster, or different versions of the same ornament being split into two clusters.

Altogether, we find it both surprising and interesting to see that no standard method enabled to perfectly solve the simple task of aligned printed patterns clustering, and believe that it demonstrates the interest of our dataset to evaluate and help design new algorithms.

4.2 Composite ornaments and element discovery

Methods As explained in Section 2, we focused on unsupervised instance segmentation approaches and evaluated the SPACE [55], AST-argmax [73], SPAIR

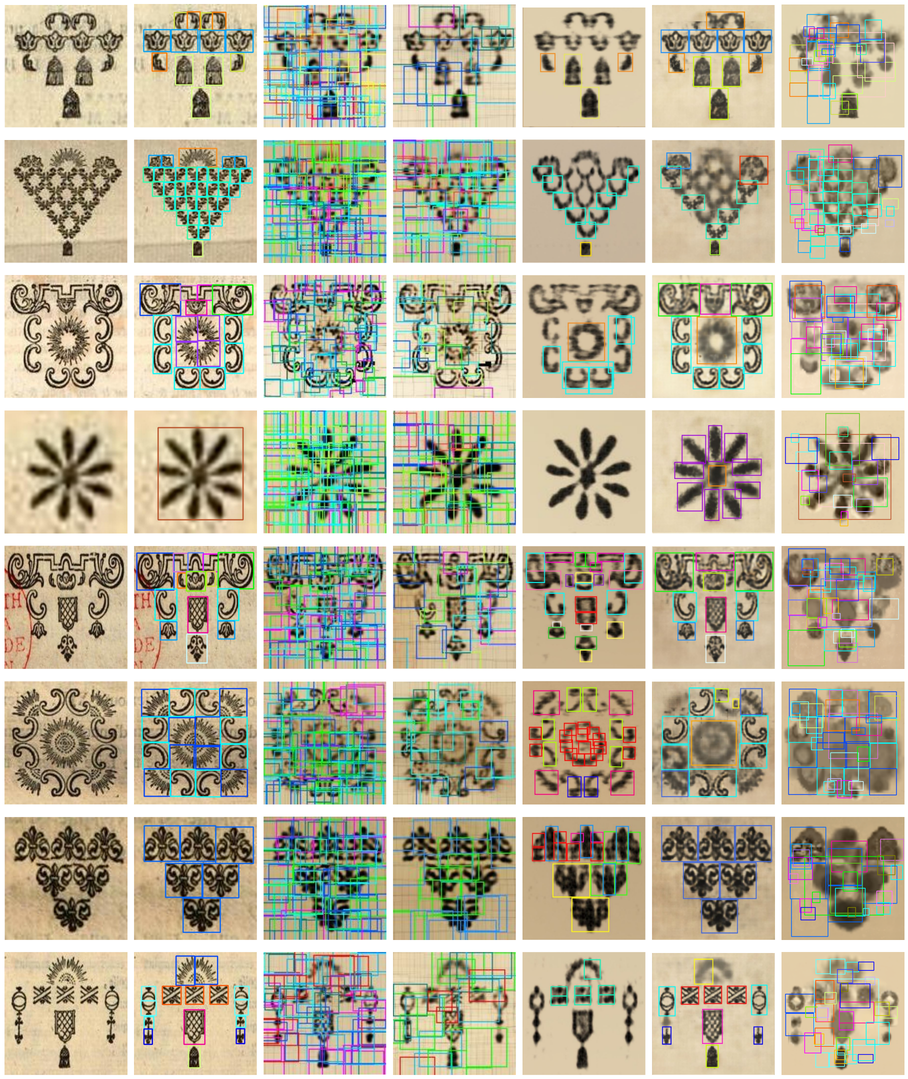
[22], GMAIR [90] and DTI-sprite approaches [62]. Only GMAIR and DTI-sprite provide categories for the different elements. Thus for all methods, we also considered performing clustering on the discovered element regions. We do so using k-means on CLIP features, which we found to perform best for clustering in the previous section.

[sayan: Through our decomposition task we highlight the complexity of discovering independent elements in the perspective of historical documents and also prove that the models prove incompetent for a task such as this, since most of the state-of-the art models have been designed to discover elements that are more geometrically conventional in nature and more or less disconnected.]

Categories		**None**		**k-means+CLIP**	
Training Data		**Real**	**Synt.**	**Real**	**Synt.**
Model	**Bkg**	AP(%) ↑	AP(%) ↑	mAP(%) ↑	mAP(%) ↑
SPAIR [22]	✗	0	0	0	0
GMAIR [90]	✗	0	0	0	0
SPACE [55]	✓	0	8.12	0	6.11
AST-argmax [73]	✓	13.58	**38.41**	13.2	**27.58**
DTI [62]	✓	0	0	0	0

Table 3: **Quantitative results for element discovery**. For each method, we report results both for models trained directly on the real composite ornaments ('Real') and pre-trained on synthetic data and fine-tuned on real data ('Synt.'). We report category agnostic average precision ('None') and mean average precision using the model categories when available ('Model') and clustering identified elements ('k-means+CLIP').

Results **[sayan: Quantitative results are reported in Table 3. The most striking fact is that all methods perform extremely poorly when trained on our real dataset, even when only the localization of the elements is evaluated (13.58% AP for the best method). We do emphasize that the absence of a background model makes it more complicated for GMAIR and SPAIR to have sufficiently strong results for the task. However, for the other models that do have a background model we see that even these models have extremely poor performance.We believe this is mainly due to two factors: first, the intrinsic complexity of the composite ornaments, where the different elements**



(a) Original (b) GT (c) SPAIR (d) GMAIR (e) SPACE (f) AST (g) DTI

Fig. 4: **Qualitative results for element discovery.** We show the dataset images (a) with their semantic ground truth bounding boxes (b) and the reconstruction and predicted semantic bounding boxed from different models (c-g). For all the methods, we show the results of the models pre-trained on the synthetic dataset and fine-tuned on the real data, and the semantic boxes obtained using K-means on CLIP features.

exist as connected components; and second, the difficult statistics of

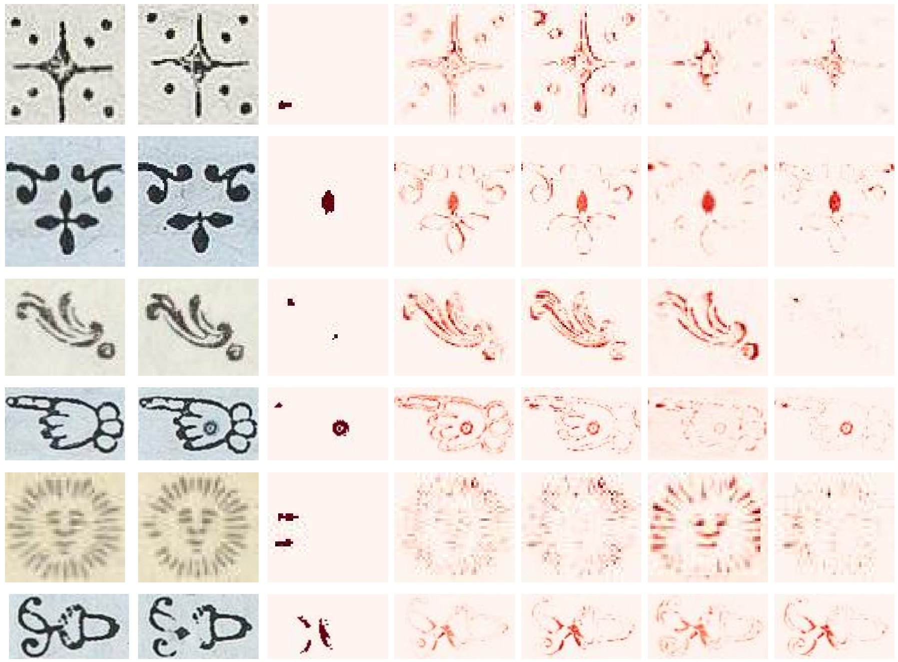
Method	Naive		VAE [50]		STAE [14]		Cong. [21]	
Category	C	CU	C	CU	C	CU	C	CU
dot1	0.0	0.0	2.5	1.8	5.2	2.5	**7.5**	**5.0**
dot2	18.0	8.9	9.7	7.3	14.7	0.7	**38.5**	**35.7**
dot3	19.8	7.5	18.2	9.3	23.4	5.7	**36.4**	**28.4**
dot4	1.6	0.0	**5.0**	**5.4**	4.7	1.2	0.0	0.0
dot5	18.2	20.0	16.7	14.5	**24.5**	**31.3**	21.8	23.5
emblem1	27.1	21.7	22.8	16.3	2.2	2.5	**32.1**	**26.3**
emblem2	28.9	22.8	37.4	26.0	**71.0**	**50.0**	62.0	47.9
emblem3	35.8	29.0	29.5	18.2	40.0	0.6	**52.5**	**42.0**
emblem4	2.3	1.0	1.5	1.1	1.7	2.2	**21.2**	**20.0**
emblem5	21.3	16.4	15.1	11.2	**28.4**	1.3	26.1	**23.6**
flower1	0.7	0.9	1.7	1.5	0.9	0.0	**15.0**	**13.6**
flower2	41.2	21.1	36.5	19.0	25.0	0.8	**50.0**	**45.9**
flower3	4.7	5.2	7.7	7.5	**26.2**	0.0	21.8	**19.3**
flower4	2.8	2.4	1.3	1.3	24.1	2.1	**28.1**	**27.9**
flower5	28.6	5.4	18.1	5.7	**43.4**	**27.3**	28.9	21.1

Method	Naive		VAE [50]		STAE [14]		Cong. [21]	
Category	C	CU	C	CU	C	CU	C	CU
interlacing1	1.2	0.8	1.1	0.6	11.9	0.0	**45.7**	**12.6**
interlacing2	**39.8**	**27.1**	30.5	20.4	27.6	15.9	36.0	26.1
interlacing3	31.3	25.0	**40.2**	**31.7**	21.9	0.0	22.8	14.6
interlacing4	32.4	17.6	39.7	18.1	35.5	16.0	**48.0**	**23.5**
interlacing5	2.8	1.3	3.1	1.8	15.6	0.3	**24.7**	**7.3**
ring1	18.5	14.8	17.7	10.3	8.3	4.3	**37.9**	**29.8**
ring2	13.5	11.1	15.9	11.8	**27.2**	**19.1**	23.3	17.5
ring3	0.6	0.6	1.4	1.2	1.9	1.5	**22.9**	**13.4**
ring4	1.5	1.5	6.3	4.5	3.9	3.8	**10.1**	**7.5**
ring5	0.5	0.8	1.4	1.5	0.0	0.0	**5.4**	**5.1**
symbol1	0.0	0.0	3.3	1.9	4.9	0.0	**7.3**	**5.4**
symbol2	8.4	3.4	13.6	2.9	7.4	6.8	**16.7**	**11.9**
symbol3	2.2	1.2	13.0	10.6	29.9	14.6	**42.7**	**19.0**
symbol4	11.9	11.3	8.3	6.9	**19.0**	5.1	16.2	**14.9**
symbol5	0.8	0.5	5.1	3.3	**22.9**	**23.5**	20.3	16.7
mIoU	13.9	9.3	14.2	9.1	19.1	8.0	**27.4**	**20.2**

Table 4: **Quantitative results for change localization.** We detail the performance of the different baselines on all of the 30 vignette categories. We also report the average value of the IoU (mIoU) on all vignettes.

the dataset, with many rare vignettes, and ornaments composed of many elements~~vignettes~~. This was the motivation to pre-train our methods on our synthetic dataset, where the elements~~vignettes~~ are sampled from the vignette dictionary uniformly and positioned randomly, making the elements easier to learn for the models. While this helps methods identify the elements on the synthetic dataset, and then boost performances when fine-tuning the models on real data, the performance remains relatively low. The qualitative results, shown in Figure 4, give more insights into the reasons for these poor performances. Methods that do not incorporate a background model (SPAIR and GMAIR) perform good reconstructions but use many elements that seem randomly localized. On the contrary, glimpse-base methods that incorporate a background model (SPACE and AST-argmax) reconstruct part of the ornaments using their background model, even if the results of AST-argmax are encouraging. It was interesting to note that, AST-argmax reproduced close to perfect reconstrcutions and discovery for some compound ornaments where as had abysmal results on other sets. Finally, we found that DTI-sprite struggled to converge on this data and produced inaccurate reconstructions.]Note that the limitations we point out are in line with the conclusion of a recent study on the potential of unsupervised object segmentation in real-world images [88].

4.3 Vignettes and unsupervised change localization



(a) Normal (b) Changed (c) GT (d) Naive (e) VAE (f) STAE (g) Cong.

Fig. 5: **Qualitative results for change localization.** For randomly selected vignettes, we show (a) an example normal vignette as well as (b) the changed vignette with (c) the corresponding ground truth change mask (GT). For each method (d-g), we show the predicted difference image.

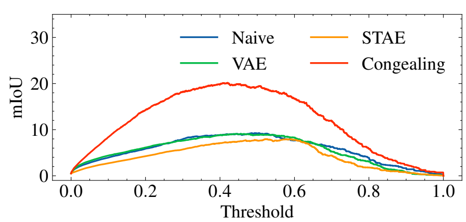


Fig. 6: **Threshold selection.** We show the mIoU of all change localization methods as a function of the threshold used to define changed pixels.

Methods. We evaluate methods that, given a test image, compute its best approximation similar to the reference images, then rely on the difference between the predicted and original image to identify the changes. The most naive approach is to approximate the test image using the average of the reference images, which provides a first baseline. A natural improvement over this method is to perform congealing [21] first on the reference images to obtain an 'aligned average', then align it to the test sample before computing the difference. Alignment is done using a color and an affine transformation. We refer to this approach as 'congealing'. We also test more advanced approaches, that rely on learning an auto-encoder on the reference samples and are thus in theory able to learn richer variations, such as the ones related to inking. We evaluate a simple variational autoencoder (VAE) [50] as well as STAE [14], a method combining a spatial transformer (ST) and a fully-connected autoencoder (AE).

Note that since all these methods compute image differences to localize the changes, computing a segmentation of the changed regions essentially requires deciding on a threshold above which we consider pixels to be changed. We computed the mIoU for each method for different thresholds, as shown in Figure 6. We found the performance to be quite stable for the different thresholds, and thus simply selected the best threshold for each method.

Results. We report the performance of the different methods in Table 4 and show qualitative comparisons in Figure 5. On average, congealing leads to the best results while the other three methods have similar performances. However, looking at the performance of each vignette and the qualitative results paints a slightly different picture. Indeed, STAE performs best in some cases, while having almost zero performance in some others. These performances irregularities seem related to the need for class-dependent threshold values in reconstruction-based methods [82].

The qualitative results also hint that morphological operations on the difference images or segmentation maps potentially joined with operations on the original images could improve the quantitative results. However, this would introduce hyper-parameters, that are unlikely to generalize beyond a specific dataset.

5 Conclusion

We have introduced a rich and historically meaningful dataset of book ornaments, with annotations and metrics for three unsupervised tasks that are of interest to book historians. We found that despite the apparent simplicity of the printed patterns, complex deep learning methods currently fail to provide satisfying results and in many cases were outperformed by simple approaches. We thus hope this work will have a significant impact both by stimulating the applications of computer vision methods to printed material and by changing the design and evaluation of clustering, element discovery, and change localization methods.

6 Acknowledgement

This work was funded by ANR ROIi project ANR-20-CE38-0005. S. Baltaci, E. Vincent and M. Aubry were supported by ERC project DISCOVER funded by the European Union's Horizon Europe Research and Innovation program under grant agreement No. 101076028 and ANR VHS project ANR-21-CE38-0008. We thank Silya Ounoughi, Thomas Gautrais, and Vincent Ventresque for their work in the collection and annotation of the datasets, and Ségolène Albouy, Raphaël Baena, Syrine Kalleli, Ioannis Siglidis, Gurjeet Singh, Andrea Morales and Malamatenia Vlachou for valuable feedbacks.

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14. pp. 622–637. Springer (2019) 5
2. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: IEEE conference on computer vision and pattern recognition. pp. 2911–2918 (2012) 3
3. Bahier-Porte, C., Fournel, T., Vial-Bonacci, L., Denis Emonet, R., Habrard, A., Ventresque, V., Gautrais, T.: Regions of interest to investigate after learning the use of ornaments by marc-michel rey. In: Conference Abstracts of Digital Humanities 2022, Pannel Computer Vision for the Study of Printers' Ornaments and Illustrations in European Hand-Press Books. pp. 66–67 (2022) 3
4. Baudrier, E., Busson, S., Corsini, S., Delalandre, M., Landré, J., , Morain-Nicolier, F.: Retrieval of the ornaments from the hand-press period: an overview. In: 10th International Conference on Document Analysis and Recognition. IEEE. pp. 496–500 (2009) 3
5. Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4. pp. 161–169. Springer (2019) 5
6. Bergel, G., Franklin, A., Heaney, M., Arandjelovic, R., Zisserman, A., Funke, D.: Content-based image recognition on printed broadside ballads: The bodleian libraries' imagematch tool. In: Proceedings of the IFLA World Library and Information Congress (2013) 3
7. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9592–9600 (2019) 5, 6
8. Bigün, J., Bhattacharjee, S., Michel, S.: Orientation radiograms for image retrieval: an alternative to segmentation. In: Proceedings of the ICPR. pp. 346–350 (1996) 3
9. Bodleian ballads: database of woodcuts, http://balladsblog.bodleian.ox.ac.uk/blog/1069 3
10. Burgess, C.P., Matthey, L., Watters, N., Kabra, R., Higgins, I., Botvinick, M., Lerchner, A.: MONet: Unsupervised Scene Decomposition and Representation. arXiv:1901.11390 [cs, stat] (Jan 2019) 4
11. Cao, L., Fei-Fei, L.: Spatially coherent latent topic model for concurrent object segmentation and classification. In: ICCV (2007) 4
12. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep Clustering for Unsupervised Learning of Visual Features. In: arXiv:1807.05520 [Cs] (ECCV 2018) 4
13. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021) 9
14. Chaki, S., Steinlin, S., Emonet, R., Fournel, T.: One-to-many pattern comparison combining fully-connected autoencoder with spatial transformer for ornament investigation. https://doi.org/10.21203/rs.3.rs-3573134/v1 (2023) 6, 13, 15

15. Cho, M., Kwak, S., Schmid, C., Ponce, J.: Unsupervised Object Discovery and Localization in the Wild. In: CVPR (2015) 4
16. Chung, J.S., Arandjelovic, R., Bergel, G., Franklin, A., Zisserman, A.: Representations of art collections. In: Workshop on Computer Vision for Art Analysis (Visart), ECCV (2014) 3
17. Compositor: database of ornaments, https://compositor.bham.ac.uk 3
18. Corsini, S.: La preuve par les fleurons: analyse comparée du matériel ornemental des imprimeurs suisses romands, 1775-1785. In: Centre international d'étude du XVIIIe siècle (1999) 3
19. Corsini, S.: « passe-partout : banque internationale d'ornements d'imprimerie. Bulletin des bibliothèques de France **5**, 73 (2001) 3
20. Corsini, S.: Vers un corpus des ornements typographiques lausannois du xviiie siècle : problèmes de définition et de méthode. In: Ornementation typographique et bibliographie historique. vol. Mons et Bruxelles, Van Balberghe, pp. 139–158 (1988) 1
21. Cox, M., Sridharan, S., Lucey, S., Cohn, J.: Least squares congealing for unsupervised alignment of images. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2008). https://doi.org/10.1109/CVPR.2008.4587573 5, 13, 14
22. Crawford, E., Pineau, J.: Spatially invariant unsupervised object detection with convolutional neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 3412–3420 (2019) 5, 11
23. Dutta, A., Bergel, G., Zisserman, A.: Visual analysis of chapbooks printed in scotland. In: The 6th International Workshop on Historical Document Imaging and Processing. pp. 67–72 (2021) 3, 6
24. Engelcke, M., Kosiorek, A.R., Jones, O.P., Posner, I.: GENESIS: Generative Scene Inference and Sampling with Object-Centric Latent Representations. ICLR (2020) 4
25. Engelcke, M., Parker Jones, O., Posner, I.: Genesis-v2: Inferring unordered object representations without iterative refinement. Advances in Neural Information Processing Systems **34**, 8085–8094 (2021) 4
26. Enschedé, J.: Proef van letteren, welke gegooten worden in de nieuwe haerlemsche lettergietery van j. enschedé (1768), https://gallica.bnf.fr/ark:/12148/bpt6k328783b 7, 8
27. Eslami, S., Heess, N., Weber, T., Tassa, Y., Szepesvari, D., Hinton, G.E., et al.: Attend, infer, repeat: Fast scene understanding with generative models. Advances in neural information processing systems **29** (2016) 5
28. Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J.M., Zisserman, A.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010), http://dblp.uni-trier.de/db/journals/ijcv/ijcv88.html 8
29. Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S.: Machine learning for cultural heritage: A survey. Pattern Recognition Letters **133**, 102–108 (2020). https://doi.org/https://doi.org/10.1016/j.patrec.2020.02.017, https://www.sciencedirect.com/science/article/pii/S0167865520300532 4
30. Fleuron: database of printing ornaments, https://db-prod-bcul.unil.ch/ornements/scripts/index.html 3
31. Fournier, P.S.: Les caracte'res de l'imprimerie (1764), https://gallica.bnf.fr/ark:/12148/bpt6k15021752/ 8
32. Frey, B., Jojic, N.: Transformation-invariant clustering using the em algorithm. Pattern Analysis and Machine Intelligence, IEEE Transactions on **25**, 1– 17 (02 2003). https://doi.org/10.1109/TPAMI.2003.1159942 4

33. Frey, B.J., Jojic, N.: Fast, large-scale transformation-invariant clustering. In: Dietterich, T., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems. vol. 14. MIT Press (2001), https://proceedings.neurips.cc/paper_files/paper/2001/file/95f6870ff3dcd442254e334a9033d349-Paper.pdf 4
34. Goyal, K., Dyer, C., Warren, C.N., G'Sell, M.G., Berg-Kirkpatrick, T.: A probabilistic generative model for typographical analysis of early modern printing. In: Annual Meeting of the Association for Computational Linguistics (2020), https://api.semanticscholar.org/CorpusID:218486915 4
35. Goyette, N., Jodoin, P.M., Porikli, F., Konrad, J., Ishwar, P.: Changedetection. net: A new change detection benchmark dataset. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops. pp. 1–8. IEEE (2012) 6
36. Grauman, K., Darrell, T.: Unsupervised Learning of Categories from Sets of Partially Matching Image Features. In: CVPR (2006) 4
37. Greff, K., Kaufman, R.L., Kabra, R., Watters, N., Burgess, C., Zoran, D., Matthey, L., Botvinick, M., Lerchner, A.: Multi-Object Representation Learning with Iterative Variational Inference. In: ICML (2019) 5, 7
38. Guo, X., Gao, L., Liu, X., Yin, J.: Improved deep embedded clustering with local structure preservation. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17. pp. 1753–1759 (2017). https://doi.org/10.24963/ijcai.2017/243, https://doi.org/10.24963/ijcai.2017/243 4
39. He, S., Samara, P., Burgers, J., Schomaker, L.: A multiple-label guided clustering algorithm for historical document dating and localization. IEEE Transactions on Image Processing **25**, 5252–5265 (2016), https://api.semanticscholar.org/CorpusID:16772542 4
40. Hsu, C.C., Lin, C.W.: CNN-based joint clustering and representation learning with feature drift compensation for large-scale image data. IEEE Transactions on Multimedia **20**(2), 421–429 (2017) 4
41. Hu, W., Miyato, T., Tokui, S., Matsumoto, E., Sugiyama, M.: Learning discrete representations via information maximizing self-augmented training. In: International conference on machine learning. pp. 1558–1567. PMLR (2017) 4
42. Huang, D., Chen, D., Chen, X., Wang, C., Lai, J.: Deepclue: Enhanced image clustering via multi-layer ensembles in deep neural networks. ArXiv **abs/2206.00359** (2022), https://api.semanticscholar.org/CorpusID:249240539 4
43. Jaderberg, M., Simonyan, K., Zisserman, A.: Spatial Transformer Networks. In: Advances in Neural Information Processing Systems. vol. 28 (2015) 6
44. Ji, X., Henriques, J.F., Vedaldi, A.: Invariant information clustering for unsupervised image classification and segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 9865–9874 (2019) 4, 9
45. Jiang, J., Ahn, S.: Generative neurosymbolic machines. Advances in Neural Information Processing Systems **33**, 12572–12582 (2020) 5
46. Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Zitnick, C.L., Girshick, R.: CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In: CVPR (2017) 5, 7
47. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image cosegmentation. In: CVPR (2010) 4
48. Kabra, R., Burgess, C., Matthey, L., Kaufman, R.L., Greff, K., Reynolds, M., Lerchner, A.: Multi-object datasets. https://github.com/deepmind/multi_object_datasets/ (2019) 5, 7

49. Karazija, L., Laina, I., Rupprecht, C.: Clevrtex: A texture-rich benchmark for unsupervised multi-object segmentation. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2) (2021) 4, 5, 7
50. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (2014) 5, 13, 15
51. Kosiorek, A.R., Sabour, S., Teh, Y.W., Hinton, G.E.: Stacked Capsule Autoencoders. In: Advances in Neural Information Processing Systems. vol. 23 (Dec 2019) 4
52. Kuhn, H.W.: The hungarian method for the assignment problem. Naval Research Logistics (NRL) **52** (1955), https://api.semanticscholar.org/CorpusID:9426884 7, 8
53. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2), 236–250 (2005) 5
54. Lesjak, Ž., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B., Špiclin, Ž.: A novel public mr image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. Neuroinformatics **16**, 51–63 (2018) 6
55. Lin, Z., Wu, Y.F., Peri, S.V., Sun, W., Singh, G., Deng, F., Jiang, J., Ahn, S.: Space: Unsupervised object-oriented scene representation via spatial attention and decomposition. In: International Conference on Learning Representations (2020), https://openreview.net/forum?id=rkl03ySYDH 5, 10, 11
56. Liu, W., Li, R., Zheng, M., Karanam, S., Wu, Z., Bhanu, B., Radke, R.J., Camps, O.: Towards visually explaining variational autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8642–8651 (2020) 5
57. MacQueen, J., et al.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. vol. 1, pp. 281–297. Oakland, CA, USA (1967) 3, 4, 9
58. Maguelone: database of typographical ornaments, http://maguelone.enssib.fr 3
59. McKenna, A., Mori, G.: Claude-françois simon, imprimeur, dit « poppy », contrefacteur de marc michel rey. In: https://mmrey.hypotheses.org/778 (2019) 1
60. Metaxas, I.M., Tzimiropoulos, G., Patras, I.: Divclust: Controlling diversity in deep clustering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3418–3428 (2023) 4, 9
61. Monnier, T., Groueix, T., Aubry, M.: Deep Transformation-Invariant Clustering. In: NeurIPS (Oct 2020) 3, 4, 9
62. Monnier, T., Vincent, E., Ponce, J., Aubry, M.: Unsupervised Layered Image Decomposition into Object Prototypes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8640–8650 (Apr 2021) 5, 11
63. Naumann, A., Hertlein, F., Dörr, L., Furmans, K.: Tampar: Visual tampering detection for parcel logistics in postal supply chains. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 8076–8086 (2024) 6
64. Pang, G., Shen, C., Cao, L., Hengel, A.V.D.: Deep learning for anomaly detection: A review. ACM computing surveys (CSUR) **54**(2), 1–38 (2021) 5
65. Pimentel, M.A., Clifton, D.A., Clifton, L., Tarassenko, L.: A review of novelty detection. Signal processing **99**, 215–249 (2014) 5

66. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021) 3, 9
67. Rey database: database of publishing ornaments, ROIi:https://heurist.huma-num.fr/heurist/?db=ROIi 3, 6
68. Riffaud, A., Pantin, I.: Une archéologie du livre français moderne. Droz (2011) 1
69. Rosart, J.: Epreuves des caractères qui se gravent et se fondent dans la nouvelle fonderie de jacques françois rosart (1761) 7, 8
70. Rubinstein, M., Joulin, A., Kopf, J., Liu, C.: Unsupervised Joint Object Discovery and Segmentation in Internet Images. In: CVPR (2013) 4
71. Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.R.: A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE **109**(5), 756–795 (2021) 5
72. Russell, B.C., Freeman, W.T., Efros, A.A., Sivic, J., Zisserman, A.: Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. In: CVPR (2006) 4
73. Sauvalle, B., de La Fortelle, A.: Unsupervised multi-object segmentation using attention and soft-argmax. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3267–3276 (2023) 5, 10, 11
74. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging. pp. 146–157. Springer (2017) 5
75. Shen, L., Lu, Y., Chen, H., Wei, H., Xie, D., Yue, J., Chen, R., Lv, S., Jiang, B.: S2looking: A satellite side-looking dataset for building change detection. Remote Sensing **13**(24), 5094 (2021) 6
76. Shen, X., Efros, A.A., Aubry, M.: Discovering Visual Patterns in Art Collections with Spatially-consistent Feature Learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Mar 2019) 4
77. Siméoni, O., Puy, G., Vo, H.V., Roburin, S., Gidaris, S., Bursuc, A., Pérez, P., Marlet, R., Ponce, J.: Localizing objects with self-supervised transformers and no labels. Proceedings of the British Machine Vision Conference (BMVC) (November 2021) 8
78. Sivic, J., Zisserman, A.: Video google: Efficient visual search of videos. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds) Toward Category-Level Object Recognition. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. vol. 4170, pp. 127–144 (2006) 3
79. Sivic, J., Russell, B.C., Zisserman, A., Freeman, W.T., Efros, A.A.: Unsupervised discovery of visual object class hierarchies. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008) 4
80. Smirnov, D., Gharbi, M., Fisher, M., Guizilini, V., Efros, A.A., Solomon, J.: MarioNette: Self-Supervised Sprite Learning. arXiv:2104.14553 [cs] (Apr 2021) 5
81. Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van Gool, L.: Scan: Learning to classify images without labels. In: Proceedings of the European Conference on Computer Vision (2020) 4, 9
82. Venkataramanan, S., Peng, K.C., Singh, R.V., Mahalanobis, A.: Attention guided anomaly localization in images. In: European Conference on Computer Vision. pp. 485–503. Springer (2020) 5, 15
83. Vo, H.V., Bach, F., Cho, M., Han, K., LeCun, Y., Pérez, P., Ponce, J.: Unsupervised Image Matching and Object Discovery as Optimization. In: CVPR (2019) 4

84. Wilkinson, H., Briggs, J., Gorissen, D.: Computer vision and the creation of a database of printers' ornaments. Digital Humanities Quarterly (2021) 3
85. Wilkinson, H.: 'printers' flowers as evidence in the identification of unknown printers: Two examples from 1715. In: The Library, 7th ser. vol. 14, pp. 70–79 (2013) 1
86. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International conference on machine learning. pp. 478–487. PMLR (2016) 4
87. Yang, B., Fu, X., Sidiropoulos, N.D., Hong, M.: Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In: international conference on machine learning. pp. 3861–3870. PMLR (2017) 4
88. Yang, Y., Yang, B.: Promising or elusive? unsupervised object segmentation from real-world single images. Advances in Neural Information Processing Systems **35**, 4722–4735 (2022) 5, 13
89. Zenati, H., Foo, C.S., Lecouat, B., Manek, G., Chandrasekhar, V.R.: Efficient gan-based anomaly detection. arXiv preprint arXiv:1802.06222 (2018) 5
90. Zhu, W., Shen, Y., Yu, L., Aguirre Sanchez, L.P.: Gmair: Unsupervised object detection based on spatial attention and gaussian mixture. arXiv e-prints pp. arXiv–2106 (2021) 5, 11