Semantic Hierarchical Exploration of Large Image Datasets

A. Bäuerle[1†], C. van Onzenoodt[2†], D. Jönsson[3†], and T. Ropinski[2],

[1]Sigma Computing, CA, USA
[2]Ulm University, Germany
[3]Linköping University, Sweden

Abstract

We present a method for exploring and comparing large sets of images with metadata using a hierarchical interaction approach. Browsing many images at the same time requires either a large screen space or an abundance of scrolling interaction. We address this problem by projecting the images onto a two-dimensional Cartesian coordinate system by combining the latent space of vision neural networks and dimensionality reduction techniques. To alleviate overdraw of the images, we integrate a hierarchical layout and navigation, where each group of similar images is represented by the image closest to the group center. Advanced interactive analysis of images in relation to their metadata is enabled through integrated, flexible filtering based on expressions. Furthermore, groups of images can be compared through selection and automated aggregated metadata visualization. We showcase our method in three case studies involving the domains of photography, machine learning, and medical imaging.

CCS Concepts

• ***Human-centered computing*** → ***Graphical user interfaces;*** *Web-based interaction; Visual analytics;*

1. Introduction

Current photo libraries can be explored by date, location, or specific metadata attributes. While these exploration approaches can be highly useful in many scenarios, semantics-based image exploration can help users find images based on their content. However, semantics-based image exploration is still an underexplored area of research [BHA*23] and often comes with significant downsides, such as overdraw or an overwhelming number of images to look at. To address this gap, we present Hierarchical Image Explorer (HIE), a semantics-based hierarchical image exploration approach.

HIE is inspired by the map-based exploration that many image libraries provide, e.g., Apple's app Photos [Inc23]. Map-based visualizations make use of the fact that, in many cases, the user knows roughly where the image of interest was taken and can then refine their search when they get closer to the true location. However, in contrast to these visual exploration tools, HIE uses semantic information embedded on a two-dimensional canvas instead of geospatial points for location assignment. Besides the layout of images, photo exploration tools also need to enable browsing large image catalogs. HIE addresses this scalability problem through an overview first, details on demand visualization. As such, only representative images of a region in the projection are displayed on a coarse level. Users can then zoom into regions of interest to inspect the content in this area in more detail. Altogether, HIE provides a hierarchical map of images, where high-level orientation helps refine the semantic exploration while zooming in. Our contributions to the field of image data set visualization are as follows:

- A hierarchical structure based solely on image embeddings that enables a multi-level semantic exploration of image data. Image groups for different hierarchy levels are automatically created, each visualized using a group representative. The structure we propose is the backbone to an overview first, details on demand mechanism for semantic image exploration.
- Design and implementation of HIE, an embedding-based, hierarchical visualization environment. HIE provides a data set overview and regional details based on the aforementioned hierarchical image structure. A user's image data can be loaded into HIE using our proposed data pipeline, which makes HIE adaptable for many use cases and different kinds of image data.

We demonstrate the functionality of HIE with three use cases in the areas of photo library search, machine learning data exploration, and medical education. HIE is available as open-source software on GitHub and can be explored online.

† Authors contributed equally to this research.

2. Related Work

Our work fits into the context of multimedia visualization techniques [ZW14]. Here, our approach fits best into the *similarity space* techniques, albeit with several augmentations to mitigate this technique's limitations.

Xie et al. implemented a semantics-based image exploration approach based on image captions [XCZ*18]. In contrast to our work, they rely heavily on extracted keywords, which might not always be available, and likely don't always represent the image semantics of interest. Consequently, the hierarchical exploration approach Xie et al. present is based on extracted keywords rather than image groups based on any semantic representation, as is the case for HIE. To tackle the problem of overdraw that embedding techniques typically have, methods such as Isomatch [FDH*15] and overlap removal [HMJE*19] have been proposed for realignment strategies, albeit without a hierarchical structure. At the same time, most research around hierarchical exploration is focused on clustering rather than embedding techniques [Tek22]. ExplorerTree is an exception to this, as they present a hierarchical exploration technique of embeddings [MJEP*21], but without packing, representatives, and many of HIE's image exploration capabilities.

With the rise in attention to machine learning visualization, it is natural that there have been various approaches towards visualizing large image data sets [YCY*21]. One example of such a visualization system is Dendromap [BHA*23], which in turn takes inspiration from PhotoMesa [Bed01] and CAT [GMIL08]. Dendromap uses a tree-based visualization approach based on a hierarchical clustering method. In contrast, HIE is solely designed to represent semantic embedding information without the assumptions or parameter settings that clustering techniques require. There also exist ML visualization interfaces that employ embedding methods [STN*16]. However, these methods suffer from a scalability problem, where the visualization either suffers from significant overdraw [NW08] or grows the canvas based on the number of images, making orientation difficult [ZXSR21]. OODAnalyzer also provides a hierarchical exploration approach [CYL*20], but their visualization is tailored to out-of-distribution analysis and does not support the more generic image exploration use case.

3. Hierarchical Image Explorer

HIE is built on a flexible data pipeline that allows users to integrate their own image data sets with custom metadata. The hierarchical exploration approach provides an overview at a coarse level while zooming into specific regions reveals more images in the semantic context of interest. HIE further provides means to select, compare, filter, and highlight data.

3.1. Data Pipeline

HIE expects a dataset to contain at least a unique image identifier and a file path in an Apache Arrow table [Fou23]. Additionally, there needs to be a data table that describes the projected location of an image by mapping its identifier to a two-dimensional position. We provide means to automatically generate this mapping to make data set creation as easy as possible. At the moment, we provide three different options for generating this mapping:

Figure 1: Coarsest honeycomb grid level with semantically embedded images of the flowers data set [Tea19]. Daisys are left of roses, while outliers can be spotted on the far right.

Embedding-based. An ML model creates a semantic embedding of the image. This semantic embedding is then projected onto a 2D canvas to obtain x and y positions using UMAP [MHSG18]. This is the default method and works best for semantic embedding.

Pixel-based. Raw pixel values of the image are used to project images onto a 2D canvas. This method acts as a fallback and only works well if images can be well categorized by their color values.

Vector-based. Users can optionally provide a custom vector for projection together with the data set. This method makes our pipeline highly customizable.

HIE loads data and projection information based on a configuration file. This way, HIE can be loaded with different projections for the same data set to compare mapping methods. The pipeline is built to scale to large image data sets (e.g., ImageNet [RDS*15] with ≈1.3M images) by streaming embeddings to disk. If necessary, the 2D projection (e.g., UMAP) can be executed on a subset of the images and then projects the entire image data set using the trained projection model. In the same way, an existing projection can be used to add new images to the data set without having to recreate the projection.

3.2. Hierarchical Exploration

To facilitate HIE's hierarchical exploration approach, embedded image positions are discretized to a regular honeycomb grid as shown in Figure 1. Images are assigned to honeycomb cells based on their projected position. We chose a honeycomb grid rather than a square design to allow for dense packing while optimizing the distance of an image to its grid cell center. One can think of this as a tradeoff between squares, which perfectly pack, but include points far from the cell center (i.e., in corners), and circles, which minimize the distance between points and their cell center but don't pack optimally.

If only one image is part of a honeycomb cell, we directly display it in its original embedding position on the two-dimensional canvas. On the other hand, if a grid cell contains more than one image, we visualize this group as a honeycomb glyph. As shown in Figure 3,

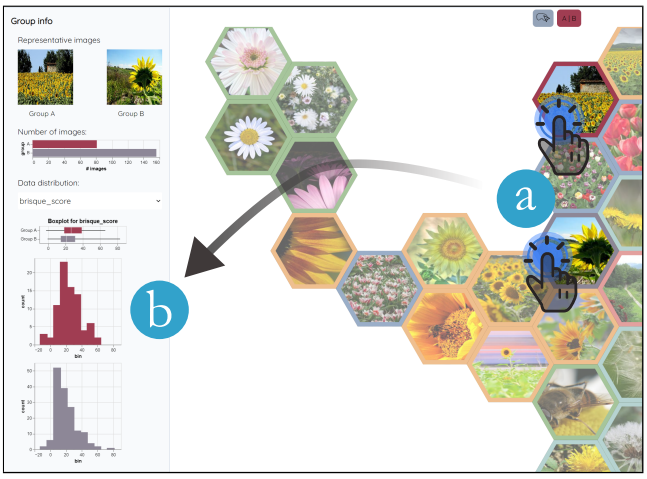

Figure 2: Illustration of image group comparison by (a) selecting one or more items and (b) inspecting their aggregated metadata. Images are the flowers data set [Tea19].

individual images are clearly visually distinct from image groups. Grid cells have a colored border that is used to surface metadata of the image group represents. Furthermore, a representative image is displayed in the honeycomb cells to provide an overview and facilitate orientation when interacting with the visualization. To fill the honeycomb, the image is proportionally scaled into the center of the image until the short side of the image fills the cell. Representative images are chosen based on their proximity to the center of the honeycomb cell. Hovering over a honeycomb enlarges it so the representative can be inspected in detail. One can set the initial resolution of the honeycomb grid used for image aggregation to accommodate different screen sizes.

If the user is interested in a particular area of the semantic grid, it is possible to zoom into the visualization. Upon zooming, the honeycombs are subdivided so that the area of interest is displayed with a higher resolution (cf. Figure 4). This approach eliminates overdraw in the visualization, reduces visualization complexity through aggregation, and provides efficient packing. To provide a global orientation when zoomed in, HIE includes a mini-map, which is commonly used in video games to provide map orientation.

We use Arrow's columnar data format for an effective computation and rendering pipeline of large image data sets. Using Arrow's filtering operations, we only load data that is currently in the viewport. Furthermore, we precompute the assignments of images to multiple honeycomb-grid levels to further reduce necessary calculations when zooming. As such, we are able to render and explore over one million images in HIE.

3.3. Metadata-based Refinement

HIE automatically integrates metadata provided through the dataset table in three ways: comparison charts, coloring, and filters. As such, HIE not only allows for semantic exploration but also metadata-based interaction and analysis.

Data set	# images train / test split	Model
Flowers [Tea19]	3.7 k	VGG16
ImageNet [RDS*15]	1.3 M / 50 k	VGG16
Malaria [RAP*18]	24.8 k / 2.8 k	Xception (finetuned)

Table 1: Overview of data sets and methods used in the use cases.

Comparing Image Groups. HIE supports selections through clicks on images or groups. Furthermore, multiple images or groups can be selected using a lasso selection tool. Upon selection, the system summarizes the selected images as shown in Figure 2. This summary includes a representative image at the top, the number of images selected below that, and metadata distribution charts at the bottom. Users can additionally select which metadata field to display distribution charts for. Metadata is automatically visualized based on its variable type. A combination of a boxplot and a histogram is displayed for quantitative data, while a bar chart is used for qualitative data. Through a toggle, users can even select two distinct groups of images. If two groups are selected, these are color-coded and displayed to be compared in this overview.

Coloring. Outlines of image group honeycombs are colored either using a categorical color scale for qualitative metadata fields or a sequential color scale for quantitative metadata fields. Users can select the metadata field to color by to tune the highlighting of image groups to their specific analysis needs.

Filtering Data. Apart from comparison charts and colored outlines, metadata fields can also be used to filter the data at hand. We use Arquero to support flexible filtering based on any metadata field of the Arrow columnar data. Such filters could include simple expressions, for example, if a column holds a given value to complex expressions involving a set of columns. Filters can also be combined using logical *and* and *or* connections. As such, users can enhance their exploration through powerful filtering methods.

4. Use Cases

To demonstrate how HIE can be used in practice, we describe three use cases in which large-scale semantic image exploration could be utilized. These use cases describe photo library search, machine learning data exploration, and medical education. The first two use cases are based on data embedded with the VGG16 model [SZ14] (last fully connected layer, fc2). This model was trained on data from ImageNet [DDS*09]. Data for the third use case is based on embeddings from the Xception model [Cho17] finetuned on the malaria [RAP*18] data set using 80% of the data and ten epochs. A summary of the data sets and models used can be seen in Table 1.

4.1. Photo Library Search

Paul is an avid nature photographer with an extensive flower photo library. He knows of applications with which he can search his library by location or date. Still, he has taken images of different flowers at the same location and time, so he is more interested in exploring the images based on their semantic content. Therefore, he

loads his photos into HIE, where they get semantically embedded using a machine-learning model.

HIE immediately provides an overview of the photos organized by their visual appearance, e.g., daisies to the left, sunflowers in the middle, and roses to the right (cf. Figure 1). He notices that close-ups of flowers are separated from fields of flowers and therefore selects images from both of these groups to inspect which ones look best, as shown in Figure 2. Paul's metadata includes brisque scores [MMB11], an image quality metric that aids Paul in determining which images are of the highest quality. By inspecting the distribution of brisque scores as shown in Figure 2, he sees that images in Group A (fields of flowers) generally are of higher image quality compared to Group B (individual flowers). Therefore, he zooms into group A and further explores the images until he finds the image he enjoys most.

4.2. Machine Learning Data Exploration

Mary is a machine learning engineer at a large software company. Her task is to investigate specific failure modes of their models. In this case, she uses the ImageNet data set with labels as well as classification results of both the VGG16 and the InceptionV3 models. She wants to investigate typical misclassifications in her data and is especially interested in patterns that might appear for multiple models.

Mary loads her data into HIE and adds a filter (label $\neq$ prediction) to show only misclassified images. From the overview, she cannot discern any particular patterns. She, therefore, zooms into an area of cars for a more detailed investigation; see Figure 3. She notices that many of the misclassified images are either sports cars or close-ups of the front of the car. By investigating the labels and the predictions of the images, she finds out that *convertible* and *sports car* are sometimes mixed up. Another typical mixup she notices is between the labels *grill* and types of cars for images showing the front of a car. She reports back to the quality assurance group to establish labeling rules for these types of images, which can then be used to improve the performance of their models.

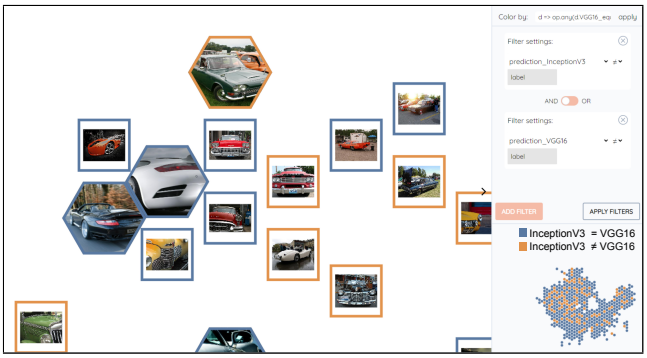

Figure 3: Zoomed-in view of cars in the ImageNet [RDS*15] data set. Multiple filter expressions (right) enable the display of misclassified images by both the InceptionV3 and VGG16 models.

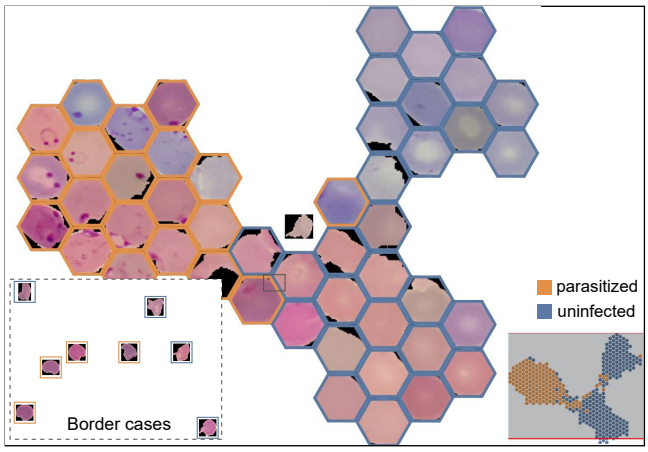

Figure 4: Inspection of blood smear slide images of malaria cases semantically embedded into a space separating parasitized and uninfected cells. The inset shows a zoomed-in view to a finer hierarchy level that allows for inspecting individual malaria cases.

4.3. Medical Education

Alva is a medical educator specializing in malaria cases. They are performing a training session about which malaria cases are difficult to recognize. These cases require special attention from medical professionals. They use a data set with images containing parasitized and uninfected cells [RAP*18]. Alva loads this data set into HIE using an AI model trained on detecting malaria.

Alva explains that the color coding (see Figure 4) refers to parasitized and uninfected cells. They then show typical cases that are easy to detect (left and right parts in Figure 4). Alva notes that the two right branches are mainly separated due to color differences, while the left branch has the typical parasitized indicators. Then, they zoom into the central part of the embedding to demonstrate border cases that are difficult to discern. The training session ends with a zoomed-in view of border case cell images (Figure 4, inset) and a discussion about how to differentiate between them.

5. Conclusion

In this paper, we present HIE, an interface for the hierarchical exploration of large image datasets. HIE is fully semantics-based and includes representatives for different levels of aggregation. We implemented a visualization interface that employs an overview first, details on demand image exploration approach. HIE incorporates various filtering and comparison mechanisms to enrich the exploration experience. To inspect image data in HIE, our flexible data pipeline that scales to large data sets, utilizes state-of-the-art embedding techniques, and integrates with available metadata can be used. HIE is available as open source software on GitHub and can be experimented with online. We demonstrated HIE's usability through three use cases and hope that HIE can help practitioners with their image exploration needs in the field.

Acknowledgements This project was supported by the Ynnerman KAW Scholar grant.

References

[Bed01] BEDERSON B. B.: Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (2001), pp. 71–80. 2

[BHA*23] BERTUCCI D., HAMID M. M., ANAND Y., RUANGROTSAKUN A., TABATABAI D., PEREZ M., KAHNG M.: Dendromap: Visual exploration of large-scale image datasets for machine learning with treemaps. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2023), 320–330. doi:10.1109/TVCG.2022.3209425. 1, 2

[Cho17] CHOLLET F.: Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1251–1258. 3

[CYL*20] CHEN C., YUAN J., LU Y., LIU Y., SU H., YUAN S., LIU S.: Oodanalyzer: Interactive analysis of out-of-distribution samples. *IEEE transactions on visualization and computer graphics 27*, 7 (2020), 3335–3349. 2

[DDS*09] DENG J., DONG W., SOCHER R., LI L.-J., LI K., FEI-FEI L.: Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (2009), Ieee, pp. 248–255. 3

[FDH*15] FRIED O., DIVERDI S., HALBER M., SIZIKOVA E., FINKELSTEIN A.: Isomatch: Creating informative grid layouts. In *Computer graphics forum* (2015), vol. 34, Wiley Online Library, pp. 155–166. 2

[Fou23] FOUNDATION T. A. S.: Apache arrow: A cross-language platform for in-memory analytics, 2023. URL: https://arrow.apache.org. 2

[GMIL08] GOMI A., MIYAZAKI R., ITOH T., LI J.: Cat: A hierarchical image browser using a rectangle packing technique. In *2008 12th International Conference Information Visualisation* (2008), IEEE, pp. 82–87. 2

[HMJE*19] HILASACA G. M., MARCÍLIO-JR W. E., ELER D. M., MARTINS R. M., PAULOVICH F. V.: Overlap removal of dimensionality reduction scatterplot layouts. *arXiv preprint arXiv:1903.06262* (2019). 2

[Inc23] INC. A.: Apple photos: Browse photos by location on iphone, 2023. URL: https://support.apple.com/guide/iphone/browse-photos-by-location-iph390138909/ios. 1

[MHSG18] MCINNES L., HEALY J., SAUL N., GROSSBERGER L.: Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software 3*, 29 (2018), 861. 2

[MJEP*21] MARCÍLIO-JR W. E., ELER D. M., PAULOVICH F. V., RODRIGUES-JR J. F., ARTERO A. O.: Explorertree: a focus+ context exploration approach for 2d embeddings. *Big Data Research 25* (2021), 100239. 2

[MMB11] MITTAL A., MOORTHY A. K., BOVIK A. C.: Blind/referenceless image spatial quality evaluator. In *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)* (2011), IEEE, pp. 723–727. 4

[NW08] NGUYEN G. P., WORRING M.: Interactive access to large image collections using similarity-based visualization. *Journal of Visual Languages & Computing 19*, 2 (2008), 203–224. 2

[RAP*18] RAJARAMAN S., ANTANI S. K., POOSTCHI M., SILAMUT K., HOSSAIN M. A., MAUDE R. J., JAEGER S., THOMA G. R.: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ 6* (2018), e4568. 3, 4

[RDS*15] RUSSAKOVSKY O., DENG J., SU H., KRAUSE J., SATHEESH S., MA S., HUANG Z., KARPATHY A., KHOSLA A., BERNSTEIN M., BERG A. C., FEI-FEI L.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV) 115*, 3 (2015), 211–252. doi:10.1007/s11263-015-0816-y. 2, 3, 4

[STN*16] SMILKOV D., THORAT N., NICHOLSON C., REIF E., VIÉGAS F. B., WATTENBERG M.: Embedding projector: Interactive visualization and interpretation of embeddings. *arXiv preprint arXiv:1611.05469* (2016). 2

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014). 3

[Tea19] TEAM T. T.: Flowers, jan 2019. URL: http://download.tensorflow.org/example_images/flower_photos.tgz. 2, 3

[Tek22] TEKLI J.: An overview of cluster-based image search result organization: background, techniques, and ongoing challenges. *Knowledge and Information Systems 64*, 3 (2022), 589–642. 2

[XCZ*18] XIE X., CAI X., ZHOU J., CAO N., WU Y.: A semantic-based method for visualizing large image collections. *IEEE transactions on visualization and computer graphics 25*, 7 (2018), 2362–2377. 2

[YCY*21] YUAN J., CHEN C., YANG W., LIU M., XIA J., LIU S.: A survey of visual analytics techniques for machine learning. *Computational Visual Media 7* (2021), 3–36. 2

[ZW14] ZAHÁLKA J., WORRING M.: Towards interactive, intelligent, and integrated multimedia analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 3–12. doi:10.1109/VAST.2014.7042476. 2

[ZXSR21] ZHAO Z., XU P., SCHEIDEGGER C., REN L.: Human-in-the-loop extraction of interpretable concepts in deep learning models. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2021), 780–790. 2