

VISUALOVERLOAD

Probing Visual Understanding of VLMs in *Really* Dense Scenes

Supplementary Material

Overview

A Quantifying Image Complexity	1
B Discussion of Domain Shift	1
C Benchmark Prompts	2
D Ablation of Resolution	2
E Language Bias Detection	2
F Performance with Advanced Prompting	3
G Embedding Space of Benchmark Questions	3
H Datasheet	5
I. Image References	8

A. Quantifying Image Complexity

Measuring visual density or complexity remains a non-trivial challenge for standard metrics [15, 36, 42, 44]. To obtain complexity measurements of our dataset, we chose to quantify complexity through relative, pairwise comparisons across the full set of 150 images. We automate this process using Qwen3.5-27B (thinking) [58] to minimize human bias, sampling responses at temperature 1.0 ($p = 0.95, k = 20, penalty = 1.0$). The model was provided with the following prompt:

Prompt for Image Complexity (Qwen3.5-27B)

```
<Image 1><Image 2> You are an expert in image analysis. Given two images, your task is to determine which image has a higher visual density and complexity. Do not attempt to identify the name of the painting. Respond with 'A' if the first image has higher complexity, 'B' if the second image has higher complexity.
```

The prompt explicitly instructs the model to ignore the identity of the artwork, encouraging a focus on visual composition rather than semantic recognition. While we observed that the model occasionally attempted identification within its reasoning traces, its final selections remained focused on density.

Ultimately, an image’s complexity score is defined as its empirical win-rate: the proportion of pairings in which the model selected it as the more complex image. Fig. 7



least complex (0.00 %)

most complex (99.3 %)

Figure 7. **Extreme ends of complexity distribution.** Shown are the least and most complex images according to our LLM-assisted measurements and their respective win-rate.

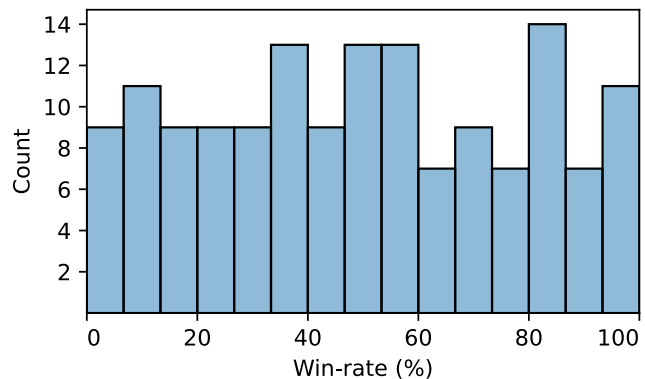


Figure 8. **Distribution of Image Complexity.** The distribution shows the win-rate over all 150 images in VisualOverload.

illustrates the images at the lowest and highest extremes of this calculated distribution (shown in Fig. 8).

B. Discussion of Domain Shift

VisualOverload consists of paintings, which will display a domain shift compared to natural images. We think of this shift as beneficial because we expect foundation models aiming for human-level perception to also be robust to stylistic abstractions.

The domain of historical paintings is a good testbed for this kind of robustness, as it offers distinct practical advantages: first, the paintings depict naturalistic scenes and objects before the time of photography. As such, they deviate from digital photos in various ways (style, color, etc.) while still maintaining the *core features* of objects, providing a good trade-off between realism and avoiding the biases of web-crawled data (center-focus, etc.). Second, we assume that this benchmark might be harder to solve than

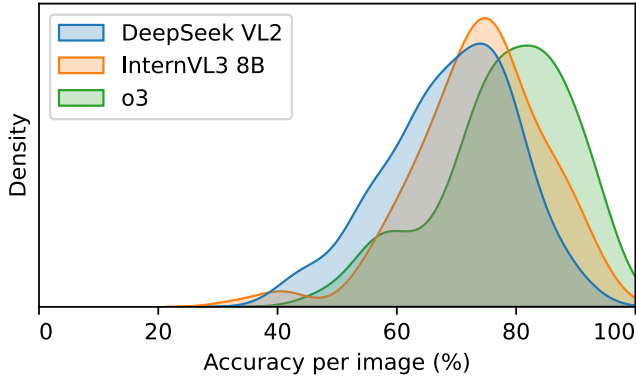


Figure 9. **Distribution of accuracy per image.** Three diverse models show distributions with clear peaks beyond average accuracy, thus proving that the challenge is caused by the content of the image and question, not the domain.

benchmarks consisting of natural images, by just crawling more data (or even generating), as the number of historical paintings is finite and not easily scalable. To improve performance, models will always, to a certain extent, have to be able to generalize. Finally, it is a domain that provides high-resolution images without depth-of-field blur (ensuring details are recognizable across the full scene) and provides a clean copyright status.

Finally, we investigate whether the domain shift may be correlated with the low performance of some models. As shown in the figure (Fig. 9), the per-image accuracy distributions for 3 distinct models lack the density near zero performance that would characterize a domain shift. Instead, they show a well-above random performance on most images, thus indicating that the performance is not tied to the domain but to the question and visual details. This is further backed by the measurements of logical consistency in Fig. 6, which show stark differences in performance for tasks differing in the required level of image understanding.

C. Benchmark Prompts

We used the following prompts in our main evaluation, depending on the question type (multiple-choice, counting, or OCR):

Default Prompt for Multiple-Choice Questions

```
{Question} Options:
A. {Option A}
B. {Option B}
...
Answer with the option's letter from the given choices directly.
```

Default Prompt for OCR Questions

```
{Question} Answer directly.
```

Default Prompt for Counting Questions

```
{Question} Answer with a number directly.
```

D. Ablation of Resolution

We distribute VisualOverload at a resolution that matches the pixels of 4K (with a few outliers). Additionally, we downsampled images to match the number of pixels of VGA (640×480 pixels), HD (1280×720 pixels), FHD (1920×1080 pixels), QHD (2560×1440 pixels), and measured task-level performances on various instances of InternVL3 models in comparison to our original resolution. The results are shown in Fig. 10.

Generally, performance improves with resolution, but at a minor rate. However, it is visible that improvements are differently correlated with tasks. Text (especially small one) is poorly compressible, and it is, thus, unsurprising to see a strong correlation between resolution and OCR performance. The opposite is modeled by scene recognition, which, for the most part, is solvable by global features that should be detectable even at extreme compression. This is backed by the lack of significant performance deviation throughout our tested resolutions. For the other tasks, we typically see an increase in performance with resolution, which seems to plateau after Full HD resolution.

This is likely not a shortcoming of our benchmark, but rather attributed to the model's architecture. By default, InternVL3 splits the input image into at most 12 patches (each 448×448 pixels) plus a thumbnail [76]. Thus, the model only supports a resolution slightly above FHD without downsampling. While it is possible to increase the number of patches, this significantly increases the inference time and memory. For instance, even for InternVL3-8B, increasing the number of patches from 12 to 40, which should be sufficient to process VisualOverload without downsampling, requires 8×40 GB GPUs, instead of just one, making such an experiment impossible for us. In theory, we, however, expect model performance to scale with resolution, assuming no downsampling. Consequently, we also expect higher performance using more patches (assuming a sufficient context window and proper training).

E. Language Bias Detection

We use Gemini 2.5 Pro with the following prompt to detect language bias:

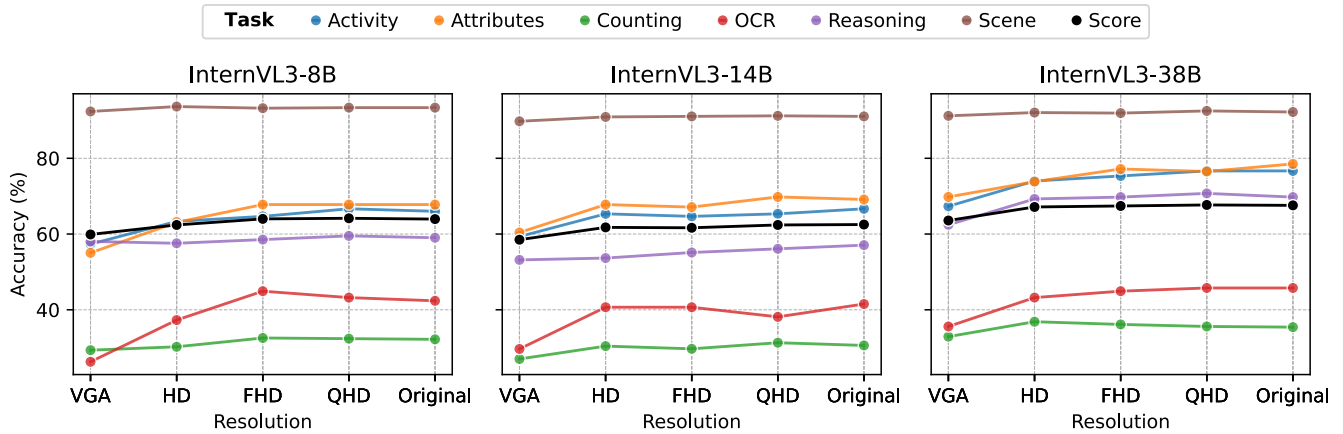


Figure 10. Resolution ablation.

Prompt for Language Bias Detection (Gemini 2.5 Pro)

Below you will find a CSV with an excerpt of questions from a visual question answering benchmark. The benchmark is supposed to be only solvable by looking at the image, however for the questions below, most models are able to guess the correct option (ground_truth). Your task is to look at each questions, the options, and ground_truth and to determine if the models were just lucky or there is some kind of shortcut or language bias. Provide an answer and rationale for each question_id.

```
question_id, question, options,
ground_truth
{CSV}
```

Tab. 2 contains the quality control results discussed in Sec. 2.1.

F. Performance with Advanced Prompting

Our evaluation in Sec. 3 utilizes simple prompts. In this section, we additionally ablate zero-shot chain-of-thought (CoT) [29, 62] on InternVL3 8B, the strongest 8B model on our benchmark, and an overall strong model. To this end, we modified the prompts as follows:

CoT Prompt for Multiple-Choice Questions

```
{Question} Options:
A. {Option A}
B. {Option B}
...
Think step by step. Answer with the option's letter from the
given choices wrapped in <answer></answer>.
```

CoT Prompt for OCR Questions

```
{Question} Think step by step. Answer with the extracted
text wrapped in <answer></answer>
```

CoT Prompt for Counting Questions

```
{Question} Think step by step. Answer with a number
wrapped in <answer></answer>
```

The results in Tab. 3 show that at least for this model, CoT decreased performance on average. However, it significantly improved performance on the hardest split and for OCR. Since CoT prompting is primarily effective in large-scale LLMs [62], we hypothesize that the tested LLM may have been too small to benefit from CoT.

G. Embedding Space of Benchmark Questions

We show a UMAP [46] reduced embedding generated by Qwen3-embedding-4B [74] of all questions (without answers) colored by task in Fig. 11. A clear separation of tasks is visible, except for the reasoning task, which overlaps with multiple other tasks as intended. The OCR questions form the most disconnected cluster.

Table 2. **Blind benchmark results.** We benchmark three models on VisualOverload without the images to measure a potential language bias.

Model	Params [B]	Activity (150)	Attributes (149)	Counting (559)	OCR (118)	Reasoning (356)	Scene (1388)	Easy (986)	Medium (1304)	Hard (430)	Total (2720)
<i>Random Chance</i>	-	25.0	25.0	0.0	0.0	25.0	25.0	24.5	16.7	3.7	16.0
<i>Consistent Chance</i>	-	25.0	25.0	0.0	0.0	42.5	50.0	47.2	26.2	4.7	27.2
InternVL3 38B [76]	38	30.0	34.9	15.6	0.8	36.6	24.2	32.5	24.5	8.1	22.8
Qwen2.5-VL 32B [5]	32	32.0	26.2	8.8	0.0	29.3	38.0	39.7	25.1	6.2	24.5
LLaVA-OV 72B [33]	72	29.3	40.3	18.1	0.8	36.1	38.6	24.6	41.1	6.5	29.2

Table 3. **Comparison with CoT prompting.**

Model	Params [B]	Activity (150)	Attributes (149)	Counting (559)	OCR (118)	Reasoning (356)	Scene (1388)	Easy (986)	Medium (1304)	Hard (430)	Total (2720)
InternVL3 38B [76]	38	76.7	78.5	35.4	45.8	69.8	92.2	99.7	81.8	7.2	67.6
+ CoT	38	74.0	69.8	34.5	50.0	62.4	91.4	98.9	77.1	14.4	65.5

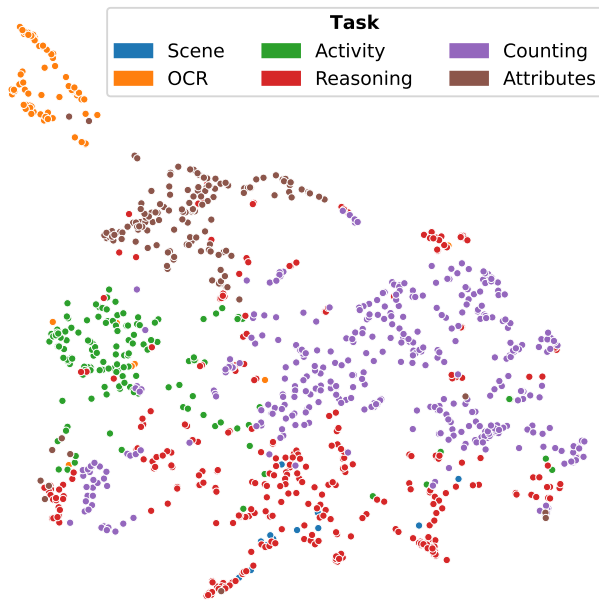


Figure 11. **Question embeddings.**

H. Datasheet

In the following, we provide a datasheet [19]. We have anonymized some entries for the review process and will update these upon release.

Motivation

For what purpose was the dataset created?

VisualOverload was created to test basic visual recognition skills of VLMs in densely populated scenes, as most prior VQA datasets often probe skills of superficial features.

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by the authors of this paper on behalf of their institutions.

Who funded the creation of the dataset?

All authors were funded by their respective institutions.

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

The dataset consists of images associated with multiple questions.

How many instances are there in total (of each type, if appropriate)?

The dataset consists of 150 images and a total of 2720 questions.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

The images are a subset of public domain artworks hosted on <https://artsandculture.google.com> filtered to display visually complex and dense scenes.

What data does each instance consist of? “Raw” data (e.g., unprocessed text or images) or features?

Each sample is a collection of the following items:

- `question_id`: Unique identifier of each question.
- `image`: A PIL JPEG image. Most of our images were resized to match the total pixel count of 4k (3840x2160 px) in different aspect ratios.
- `question`: A question about the image.
- `question_type`: Type of question. Will be one of choice (response expected to be "A", "B", "C", or "D"), counting (freeform), or ocr (freeform). You can use this information to request a suitable output format.
- `options`: This is the list of options for `question_type=choice` and empty otherwise. Please treat the options as answer options A, B, C, D (4 options) or A, B (2 options).

- `difficulty`: Meta-data about the difficulty of the question. One of easy, medium, or hard.
- `category`: Meta-data about the question task. One of activity, attributes, counting, ocr, reasoning, or scene.
- `default_prompt`: You can use this prompt to stay compliant with our results. It is a simple combination of the question and answers, with some additional output format constraints. This should work well for most models.

Further, we provide a database linking each `image` to its respective `complexity_win_rate`, as defined in Appendix A.

Is there a label or target associated with each instance?

Each question is associated with a ground-truth. This ground-truth is hidden from the public to avoid test leakage.

Is any information missing from individual instances?

We obfuscate image file names and question IDs to reduce knowledge priors.

Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)?

The samples in the dataset shall be treated independently.

Are there recommended data splits (e.g., training, development/validation, testing)?

All the samples in our dataset shall be exclusively treated as a test set. We do not provide development sets, as we consider all questions to be solvable with a basic set of skills that should be present in frontier VLMs.

Are there any errors, sources of noise, or redundancies in the dataset?

All questions and ground truths are manually annotated and, thus, may contain errors. To reduce the error rate, we double-checked all questions where multiple models provided wrong answers.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?

The dataset is self-contained.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals non-public communications)?

No.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?

The dataset contains samples that show religious beliefs, (partial) nudity, and/or injury and death.

Does the dataset relate to people?

The dataset contains artworks that may depict people.

Does the dataset identify any subpopulations (e.g., by age, gender)?

The dataset does not identify any subpopulations.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?

Some of the individuals are of historical, biblical, or mythical origin and may be identified. No living individuals can be identified from the dataset.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?

No.

Collection Process

How was the data associated with each instance acquired?

Please see Sec. 2.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

Please see Sec. 2.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

n/a.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

The dataset was collected and annotated by the authors of this paper. No crowdworkers, students, or contractors, *etc.*, were involved.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?

The images were collected between April and May 2025, and annotated and cleaned between May and August 2025.

Were any ethical review processes conducted (e.g., by an institutional review board)?

No.

Does the dataset relate to people?

The dataset contains artworks that may depict people.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

n/a.

Were the individuals in question notified about the data collection?

All depicted individuals are no longer alive.

Did the individuals in question consent to the collection and use of their data?

n/a.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?

n/a.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?

n/a.

Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?

Yes, see Sec. 2.

Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?

The raw data can be requested from the authors.

Is the software used to preprocess/clean/label the instances available?

The images were obtained using <https://github.com/lovasoa/dezoomify-rs>. All further processing scripts were developed by the authors and are not available publicly.

Uses

Has the dataset been used for any tasks already?

The dataset has been used to evaluate basic visual skills of frontier VLMs in Sec. 3.

Is there a repository that links to any or all papers or systems that use the dataset?

We will list relevant papers at <https://github.com/paulgavrikov/visualoverload>. We encourage authors to contact us to list their works.

What (other) tasks could the dataset be used for?

The dataset is primarily designed for visual question answering (VQA), but we encourage users to apply it to other tasks as desired.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

No.

Are there tasks for which the dataset should not be used?

This dataset is released exclusively for academic research and educational use. It must not be applied to purposes that could lead to harm, including surveillance, discrimination, exploitation, harassment, or the generation of misleading or offensive content. Users are expected to uphold the highest standards of research integrity and ethics, and to ensure that their work with this dataset aligns with responsible AI principles.

Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?

The dataset is publicly available.

How will the dataset be distributed (e.g., tarball on website, API, GitHub)

The dataset is distributed through HuggingFace datasets, which currently uses a PyArrow format.

When will the dataset be distributed?

The dataset is immediately distributed through: <https://huggingface.co/datasets/paulgavrikov/visualoverload>.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

The dataset is distributed under the Creative Commons Attribution-ShareAlike 4.0 International license without any further terms of use.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No.

Maintenance

Who will be supporting/hosting/maintaining the dataset?

The authors will be supporting/hosting/maintaining the dataset.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

The authors can be contacted via GitHub issues at: <https://github.com/paulgavrikov/visualoverload/issues>.

Is there an erratum?

No.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?

The dataset will not be modified to ensure comparability of results. Corrected or derived datasets will be released independently.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?

n/a.

Will older versions of the dataset continue to be supported/hosted/maintained?

The dataset will remain available as long as it continues to be hosted by the third-party platforms on which it is stored.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Users can extend/augment/build upon the dataset, but must publish their new work as a standalone derivative. We kindly request that users communicate any releases to the authors.

I. Image References

Table 4. **List of artworks.** This table contains all artworks present in VisualOverload in random order. The metadata is taken from the references with minor postprocessing by us.

Creator	Title	Date	URL
Unknown	Wood of the Philosophers	1800/1830	Link
Pieter Aertsen	The Fat Kitchen. An Allegory	1565-1575	Link
Unknown	A handscroll painting of the porcelain production process (right half)	early 19th century	Link
Avercamp, Hendrick	Enjoying the Ice	ca. 1615-1620	Link
Charles M. Russell	The Medicine Man	1908	Link
Pieter van der Heyden after Pieter Bruegel the Elder	The Big Fish Eat the Little Fish	published 1557	Link
Charles Fairfax Murray	Allegory of Good Government, after Ambrogio Lorenzetti	1873	Link
Philip Galle after Pieter Bruegel the Elder	Prudence	published 1559	Link
Pieter Bruegel the Elder, Frans Hogenberg	The Kermis at Hoboken	ca. 1559	Link
Joan Antigó, Honorat Borrassà i Francesc Vergós	Altarpiece of Saint Miquel de Castelló d'Empúries (detail)	1448	Link
Greek artist from the end of the 18th century	St. George	1798/1798	Link
Pieter Bruegel the Elder, Pieter van der Heyden, Hieronymus Cock	Anger (Ira) from The Seven Deadly Sins	1558	Link
Dirck Franchoisz Hals, Dirck van Delen	Festive Company in a Renaissance Room	1628	Link
Philips Galle, Pieter Bruegel the Elder, Hieronymus Cock	Charity (Charitas) from The Virtues	1559	Link
Unknown	A Sunday on La Grande Jatte	1884-1886	Link
Charles William Sharpe	The Death of Nelson at the Battle of Trafalgar	1806/1876	Link
Ast, Balthasar van der	Still Life with Fruit and Flowers	1620-1621	Link
School of Canaletto	St. Marks, Venice	unknown	Link
Pieter Bruegel the Elder	The Sermon of Saint John the Baptist	1566	Link
Pieter Aertsen	Market Scene	1569	Link
Thomas Matthews Rooke	Washing Sheds at Chartres	1885	Link
Severin Roesen	Still Life of Flowers and Fruit with a River Landscape in the Distance	1867	Link
Pere Mates	Final Judgment. Altarpiece of Santa Maria de Segueró (Garrotxa)	1500/1550	Link
Steen, Jan Havicksz	Villagers Merrymaking Outside an Inn	1652	Link

Hieronymus Bosch	Ecce Homo	1500	Link
Jan Steen	Beware of Luxury (“In Weelde Siet Toe”)	1663	Link
Unknown	Christ in the House of Martha and Mary	1553	Link
Ostade, Adriaen van	Peasants in an Interior	1661	Link
Pieter Bruegel the Elder	Desidia (Sloth)	1557	Link
Avercamp, Hendrick	Enjoying the Ice near a Town	ca. 1620	Link
Rijn, Rembrandt van	The Night Watch	1642	Link
Ditlev Blunck	Danish artists at the Osteria La Gensola in Rome	1837	Link
Konstantin Makovsky	A Boyar Wedding Feast	1883	Link
Unknown	The trial of the Neptune’s seamen	1807	Link
Albrecht Altdorfer	Christ taking Leave of his Mother	probably 1520	Link
Baines, Thomas	Kaffirs and Rebel Hottentotts Attacking a Wagon Train	1851/1852	Link
Unknown	Bucentaur’s return to the pier by the Palazzo Ducale	1728/1729	Link
Neer, Aert van der	Winter Landscape near a Town with Kolf Players and Horse-Drawn Sleighs	ca. 1650-1655	Link
Philip Galle after Pieter Bruegel the Elder	Faith	published 1559	Link
Francesco Hayez	Pope Urban II Preaching the First Crusade in the Square of Clermont	1835/1835	Link
Pieter Bruegel the Elder	The Fall of the Rebel Angels	1562	Link
Atelier de Paris	Psyché rapporte la laine des brebis	1650	Link
Hals, Dirck	The Fête champêtre	1627	Link
Baines, Henry	Fisherfleet Looking East	1823/1894	Link
Jan Cornelisz. Vermeyen	The Spanish Brothel	1545	Link
Marià Vayreda i Vila	Gambeto dance in Riudaura	1890	Link
Paolo De Matteis	St. Nicolas of Bari Felling a Tree Inhabited by Demons	1727/1727	Link
Philip Galle, Pieter Bruegel	The Resurrection of Christ	ca. 1562	Link
Pieter Aertsen	Market Scene	1550	Link
Pieter Aertsen	Christ with Mary and Martha	1552	Link
Philip Galle after Pieter Bruegel the Elder	The Parable of the Wise and Foolish Virgins	ca. 1560/1563	Link
Unknown	Dragon Boats at Aberdeen Hong Kong showing Careening Island	1923	Link
Pieter Bruegel the Elder, Pieter van der Heyden, Hieronymus Cock	Avarice (Avaritia), from the series The Seven Deadly Sins	1558	Link
Hendrick Avercamp	A Scene on the Ice	ca. 1625	Link
Attributed to Jan van Belcamp	The Great Picture	1646	Link

Aleksander Ivanov	The Apparition of Christ to the People (The Apparition of the Messiah)	1837-1857	Link
Hendrick Avercamp	Frozen River with Skaters	1620s	Link
Jan Rost	The Pharaoh Welcomes Joseph	1553	Link
Aertsen, Pieter	Wing of an Altarpiece with Adoration of the Magi, on the reverse is Presentation in the Temple	1560-1565	Link
Van Aachen, Hans	The Rape of Proserpine	1589	Link
Avercamp, Hendrick	Winter Landscape with Ice Skaters	ca. 1608	Link
Estevão Silva	Untitled	1887/1887	Link
Pieter Bruegel, J. Liefrinck, H. Hondius	The fat kitchen	1563	Link
Unknown	Moses descends from Mount Siniaï with the Ten Commandments	1662	Link
Unknown	British forces receiving Commissioner Keying at Canton's British Factories for conference with Sir J. F. Davies	1847	Link
Unknown	Le peuple rend les honneurs à Psyché	1650	Link
Pieter Bruegel the Elder	Peasant Wedding	1566-1569	Link
Jan Miense Molenaer	Self-Portrait with Family Members	1630/1640	Link
Philip Galle after Pieter Bruegel the Elder	Temperance	published 1559	Link
Anonymous	Christ as the Good Shepherd	1505	Link
Dirck Jacobsz Vellert	The Flood	1544	Link
Cornelis Cornelisz van Haarlem	The Golden Age (Bacchanal) or the Garden of Love	1614	Link
Aertsen, Pieter	The Adoration of the Magi	ca. 1560	Link
William Duffield	Still Life	1859	Link
Balthasar van der Ast	Still Life of Flowers, Fruit, Shells, and Insects	About 1629	Link
Albrecht Dürer	Feast of Rose Garlands	1506	Link
Jan Steen	The Dancing Couple	1663	Link
Jan Steen	The Worship of the Golden Calf	ca. 1672-1675	Link
Carl Bloch	In a Roman Osteria	1866	Link
Avercamp, Hendrick	Ice-Skating in a Village	ca. 1610	Link
After Pieter Bruegel the Elder	The Festival of Fools	after 1570	Link
Pieter van der Heyden after Pieter Bruegel the Elder	The Witch of Malleghem	published 1559	Link
Pieter van der Heyden, Pieter Bruegel the Elder, Hieronymus Cock	Patience (Patientia)	1557	Link
Francisco de Goya	El Entierro de la Sardina	1808/1812	Link

Master of Okolično	Holy Kinship	1510	Link
Maarten van Heemskerck	The Gods of the Olympus	1556	Link
Gerrit van Honthorst	Apollo and Diana	1628	Link
Pieter Bruegel the Elder	Children's Games	1560	Link
Hendrick Avercamp	Skating Scene	1620s	Link
Pieter Bruegel the Elder	The Adoration of the Magi	undated	Link
Bol, Ferdinand	Consul Titus Manlius Torquatus Orders the Beheading of his Son	1661-1663	Link
Balthasar van der Ast	Still Life with Basket of Fruit	1622	Link
Pieter Bruegel the Elder, Philips Galle, Hieronymus Cock	The Alchemist	after 1558	Link
Claesz., Pieter	Still Life with a Turkey Pie	1627	Link
Pieter Bruegel the Elder	The Dirty Bride or the Marriage of Mopsus and Nisa	1570	Link
Western artist painted from the deck of HMS Vulcan	British and French fleets in Victoria Harbour	1860	Link
Pieter Bruegel the Elder	Hunters in the Snow (Winter)	1565	Link
Philips Galle, Pieter Bruegel the Elder, Hieronymus Cock	Justice (Justicia) from The Virtues	ca. 1559–60	Link
Giulio Romano	Chamber of the Giants - Ceiling	1532-1534	Link
Pieter Bruegel the Elder, Pieter van der Heyden, Hieronymus Cock	The Descent of Christ Into Limbo	ca. 1561	Link
Hendrick Avercamp	Winter Landscape	1600/1620	Link
Arcadi Mas i Fondevila	The Corpus Christi procession	1887	Link
Unknown	Winter Scene on a Frozen Canal	1620	Link
Caroline Le Souef	Home Life of the Victorian Aborigines	1895-1895	Link
Andrej Janez Herrlein	Ljubljana Šempeter	1798	Link
Edward Roper	Gold Diggings, Ararat	1855-1860	Link
Pieter Bruegel the Elder	Massacre of the Innocents	1565-1567	Link
Lucas van Leyden	Christ presented to the people	1510	Link
Hans Bol	Goose Snatching	1560/1580	Link
Edouard Hildebrand	Menino com patos	1800/1800	Link
Pieter Bruegel the Elder	The Census at Bethlehem	1566	Link
Frans Huys	Ice Skating before the Gate of Saint George in Antwerp	1558	Link
Pieter Aertsen	A Meat Stall with the Holy Family Giving Alms	1551	Link
Philip Galle after Pieter Bruegel the Elder	Fortitude	published 1559	Link
Lucas Cranach the Elder	The Crucifixion	1506-1520	Link
Bol, Ferdinand	Aeneas Crowning Cloanthus	ca. 1661-1663	Link

Pelegrí Clavé i Roqué	Jacob receives the bloody tunic of his son Joseph	1842	Link
Aertsen, Pieter	The Egg Dance	1552	Link
Unknown	The Village Wedding	1653	Link
Joos van Craesbeeck	The Temptation of Saint Anthony	1650	Link
Pieter Bruegel the Elder	Winter Landscape with Skaters and Birds Trap	1565	Link
Caroline Le Souef	Native Fight on the Lower Goulburn River in 1842	1895-1895	Link
Joachim Wtewael	The Annunciation to the Shepherds	1606	Link
Joachim Beuckelaer	Woman Selling Vegetables	second half of 16th century	Link
Unknown	Dragon boat racing at Spring Festival	1860	Link
Karl Brullov	The Last Day of Pompeii	1830/1833	Link
Caroline Le Souef	Corroboree on the Goulburn River	1895-1895	Link
Pieter Bruegel the Elder, Hieronymus Cock, Philips Galle	Hope (Spes) from The Virtues	ca. 1559–60	Link
Steen, Jan Havicksz.	The Feast of St Nicholas	1665-1668	Link
Hans Memling	Virgin and Child with Saints Catherine of Alexandria and Barbara	early 1480s	Link
Gerard Dou	The Quack	1652	Link
Baines, Thomas	Forces under General Cathcart crossing the Orange river to attack Moshesh 1852	1854	Link
Namcheong	Whampoa's earliest mud dock, Canton	ca. 1850s	Link
Jan Fyt	Still life with parrot	ca. 1645	Link
Oldmeadow, William H.	Reffley Spring	1818	Link
Bartholomeus van Bassen, Esaias van den Velde	Renaissance Interior with Banqueters	1618/1622	Link
Gift of Mr. Antony J. Hardy	Bamboo Town and Anchorage at Whampoa Island	1860	Link
Unknown	A handscroll painting of the porcelain production process (left half)	early 19th century	Link
Pieter Brueghel, Pieter van der Heyden, Hieronymus Cock	Luxuria, uno de los siete vicios	1558	Link
Pieter Brueghel	La soberbia Serie de los siete pecados capitales	1558	Link
Pieter Bruegel the Elder	Landscape with the Fall of Icarus	undated	Link
Pieter Bruegel the Elder, Pieter van der Heyden, Hieronymus Cock	The Last Judgment	1558	Link
Hans Memling	The Annunciation	1480–89	Link
James Taylor	Panorama du Port Jackson et de la ville de Sidney	1820/1825	Link
Avercamp, Hendrick	Frolicking on a Frozen Canal in a Town	ca. 1615-1620	Link

After Pieter Bruegel the Elder	Gluttony (Gula) from The Seven Deadly Sins	1558	Link
Felip Massó i Falp	The Procession of St. Bartolomew	1884	Link
Pieter Bruegel the Elder	The Dutch Proverbs	1559	Link
Maertan Van Heemskerck	Concert of Apollo and the Muses on Mount Helicon	1565	Link
David Teniers the Younger	The Surgeon	1670s	Link