2.5 Years in Class: A Multimodal Textbook for Vision-Language Pretraining

Supplementary Material

8. Detail of Video-to-Textbook Pipeline

8.1. The Novelty of our Pipeline

- We develop a novel video-to-textbook pipeline. While it integrates some existing methods, our pipeline enables **fine-grained** knowledge extraction and **precise quality/-topic control**, allowing us to create a high-quality knowledge dataset from videos.
- Compared to prior datasets, our textbook brings notable **pretraining improvements(+5.5)**, especially **knowledge and reasoning tasks(+24)**,and in-context learning abilities.
- We **explores many practical designs** for visual/textual knowledge extraction: Which keyframe algorithm is suitable (structured-, pixel-, semantics-based)? How to filter irrelevant frames at multi-level? Does raw ASR need revision?

8.2. Copyright Statement

We strictly adhere to copyright regulations. (1) Our dataset is for academic research only. Commercial use must comply with original licenses. (2) Following most existing video dataset practices (e.g., HDVILA-100M, Panda70M), we only provide video IDs, URLs, ASR, keyframes, and timestamps, but not source videos.

8.3. Implementation Details

When synthesizing the Knowledge Taxonomy, we utilize `GPT-4o` to construct the taxonomy. When filtering video at the metadata level, `GPT-4o` is also employed to review the metadata of the searched videos. During the Video-to-ASR phase, `Whisper-large-v3` is used to convert audio into text. Then `Qwen2-72B-Instruct` is applied to refine the raw ASR transcriptions. In the video-level filtering stage, `DeepSeek-V2` and `Llama3-70B-Instruct` are used to score each ASR transcription, enabling the filtration of low-quality videos. A video is filtered out if both LLMs determine its ASR does not meet the required standards. After splitting long videos into short clips, we first use `VideoLlama2-7B` to generate a detailed caption for each video clip. Subsequently, we compute the similarity between the clip's caption and the ASR using `GTE-Qwen2-7B-Instruct`. Finally, `InternVL2` is employed to extract and filter OCR from the keyframe.

8.4. Human Evaluation

We randomly sample 100 examples from the multimodal textbook and conduct a manual quality evaluation, focusing on three key aspects: (1) image quality, (2) the connections between different images in a sample and (3) the relevance between texts and images. After manual inspection, we observe that, aside from chemistry, this batch of samples covers five domains: mathematics (31), physics (16), computer science (16), engineering (25), and earth sciences (12). It contains a total of 1,421 images, including 378 slide-style images, 214 lecture-style images, 414 demonstration animations, and 415 natural scenes. Image analysis reveals that only 7% (72 images) are highly similar, while the remaining images are related to each other but also exhibit clear distinctions. Text-image relevance analysis shows that the attached text (ASR) correctly explains the visual concepts or computational processes presented in the images, with no ambiguity or redundancy.

Algorithm 1 SSIM-Based Key Frame Extraction Algorithm

Require: Frame sequence $\{F_1, F_2, \ldots, F_N\}$, similarity threshold T
Ensure: Key frame sequence $\{K_1, K_2, \ldots\}$

```
K ← {F_1}                          ▷ Initialize key frame sequence with the first frame F_1
reference_frame ← F_1              ▷ Set the reference frame to F_1
for i = 2 to N do
    SSIM ← CalculateSSIM(reference_frame, F_i)   ▷ Calculate SSIM between reference frame and frame F_i
    if SSIM < T then
        K ← K ∪ {F_i}              ▷ If SSIM is below threshold, add frame F_i as a key frame
        reference_frame ← F_i      ▷ Update the reference frame to F_i
    end if
end for
return K                           ▷ Return the sequence of key frames
```

8.5. Constructing Pretraining Sample

After collecting 6.5M keyframes, and 750M refined ASR, and OCR tokens, we can employ various strategies to construct image-text interleaved samples for pre-training. ① Similar to a webpage-centric dataset, where each webpage is treated as a separate sample, we treat each video as an individual sample. This simple strategy maintains the semantic integrity of a video. However, it also leads to overly long contexts for most samples, as each video contains an average of 86 keyframes, far exceeding the maximum context length supported by most VLMs. ② As an alternative, we segment a single long video into multiple samples. It can flexibly segment videos based on the maximum context length supported by VLMs. ③ Besides, we directly concatenate multiple video clips i.e.,

$\langle \text{frame}_i^{k_1}, .., \text{frame}_i^{k_n}, \text{ocr}_i, \text{asr}_i \rangle$, to the maximum context length. This strategy breaks video boundaries, effectively utilizing computational resources. However, mixing multiple video clips within a single sample may adversely affect training performance. Therefore, we insert a specific token: `End of Video` at the end of each video to mitigate this.

8.6. Knowledge Taxonomy

As stated in the main text, to include richer knowledge in our textbook, we propose a hierarchical knowledge taxonomy comprising four hierarchical layers, namely *Subject* $\rightarrow$ *Course* $\rightarrow$ *Sub-course* $\rightarrow$ *Knowledge Point*. We instruct an LLM to span the knowledge taxonomy across multiple educational stages (from primary school to middle school) and diverse subjects (mathematics, physics, etc.). Lastly, we obtain a knowledge taxonomy comprising 6 subjects (mathematics, physics, chemistry, earth science, engineering, and computer science), 55 courses (Algebra, Solid Geometry,..), and 3915 knowledge points. As illustrated in Fig. 4, we plot six subjects along with their corresponding courses. Due to space constraints, we visualized the top 9 courses and their proportion. The number of knowledge points included in each course is approximately the same.

8.7. Detail of InSI-SIM

As mentioned in Sec. 4.2, we design an in-sample image similarity metric (InSI-SIM). It measures the similarity between all images within a sample. Formally, for a subset D containing M samples, each comprising L images, the in-sample image similarity is computed as follows:

$$\text{InSI-SIM}^L = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\binom{L}{2}}\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}\Big(\text{CLIP}(\text{Img}_{k,i}, \text{Img}_{k,j}) + \text{SSIM}(\text{Img}_{k,i}, \text{Img}_{k,j})\Big)/2 \tag{1}$$

where $\text{CLIP}(\text{Img}_{k,i}, \text{Img}_{k,j})$ and $\text{SSIM}(\text{Img}_{k,i}, \text{Img}_{k,j})$ represent the semantic and structural similarity scores between images i and j in sample k, respectively.

9. Details of Experiments

9.1. Data Quality Analysis

Following LAION-5B, we evaluate data quality on 8 metrics, e.g., *relevance*, *NSFW score*, *unique image proportion*, and *text perplexity*.

- Results show our dataset outperforms others on **most metrics**, except image aesthetic and text fluency.
- We analyze the **impact of different filtering strategies** on dataset quality and training results (Table 8). E.g., pixel-level keyframes cause image redundancy (Unique rate: 86→61, Acc:↓9), while ASR refinement improves text fluency (PPL: 16.8→12.7, Acc: ↑4.9).
- **Following previous datasets**, we sample 100-200 cases for **human evaluation** on image-text relevance, image quality, coherence, and knowledge density. Human inspection confirms reliable quality of ours.

9.2. Detail of Evaluation

We evaluate the pre-trained VLMs on two VQA benchmarks (TextVQA, OKVQA), a knowledge-centric benchmark (ScienceQA), and three math-related benchmarks (MathVista, MathVerse, MathVision) under few-shot settings. Following the previous works [26], we use the RICES-based few-shot prompting strategy which retrieves the k most similar samples from the training set based on the testing image feature. It should be noted that since MathVista, MathVerse, and MathVision only contain testing sets, we can not retrieve samples from their respective training sets. Consequently, for MathVista and MathVerse, we retrieve k examples from MathVision, while for MathVision, we retrieve examples from MathVista. When evaluating, we adopt the same prompt as Llava-1.5:

```
System Prompt: A chat between a human and
an artificial intelligence assistant. The
assistant gives helpful, detailed, and
polite answers to the human's questions
USER: <image>\n{example1 query}\nAnswer the
 question using a single word or phrase.
ASSISTANT: {example1 answer}</s>
USER: <image>\n{example2 query}\nAnswer the
 question using a single word or phrase.
ASSISTANT: {example2 answer}</s>
....
USER: <image>\n{testing query}\nAnswer the
question using a single word or phrase.
ASSISTANT:
```

10. Limitations

Although we already designed multiple levels of filtering, our textbook may still contain some redundant keyframes, low-quality texts, and so on. We will continue to improve the quality and knowledge density of our textbook. Besides, similar to prior multimodal models, our textbook primarily focuses on multimodal understanding and text generation for interleaved contexts. During training, the loss is not computed for image tokens. However, our textbook can also be used for omni-modal models including both understanding and generation tasks. We leave this for future work.

10.1. Examples of Multimodal Textbook

We provide several detailed examples in Figs. 5 to 10. Specifically, Fig. 5 offers a detailed explanation of the Earth's water cycle, presented through slides, photographs, and schematic diagrams. Figures 6 and 7 provide rich visualizations, including diagrams and texts, to elucidate the

Subject	#Video	Duration (h)	#Topic	#Video Clip	#Keyframe	#ASR Token	#OCR Token	#Sample
Mathematics	21.7k	4,423	725	809k	1.67M	72.5M	145M	123k
Physics	11k	3,511	530	822k	0.95M	36.7M	73.4M	119k
Chemistry	4.5k	2,643	410	234k	0.49M	15M	30M	32k
Earth Science	12k	3,670	520	640k	1.03M	40M	80M	88k
Engineering	13k	4,096	810	713k	1.15M	43.3M	86.6M	98k
Computer Science	12.8k	4,354	820	782k	1.21M	42.8M	85.5M	150k
All	**75k**	**22,697**	**3,915**	**4M**	**6.58M**	**258M**	**500M**	**610k**

Table 7. The statistics of our multimodal textbook. Topic denotes the knowledge points covered by each category of videos, which are sourced from our knowledge taxonomy.

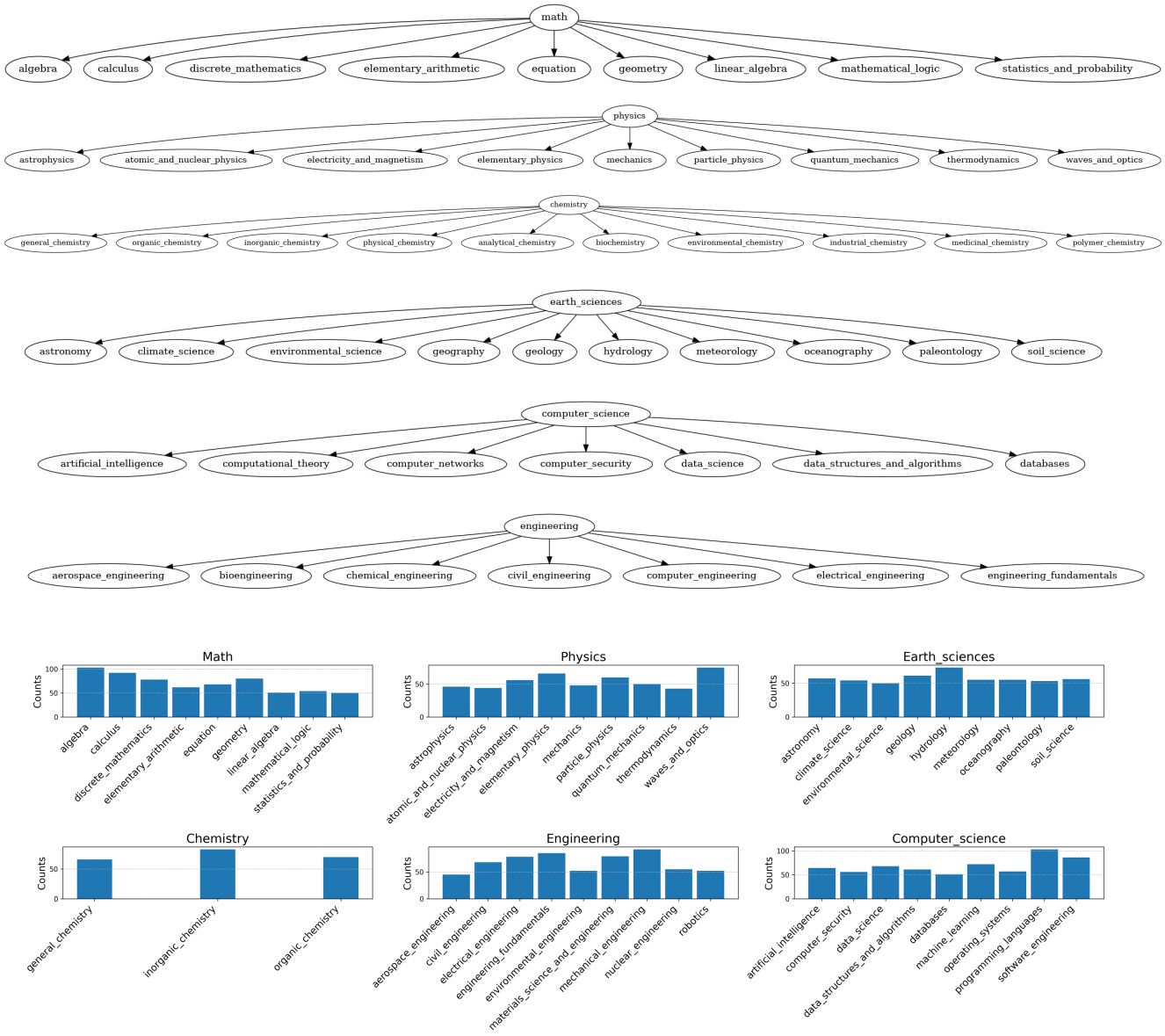


Figure 4. Top: We plot six subjects along with their corresponding sub-courses. Due to space constraints, we selectively visualized only the courses with the highest proportions. Bottom: We count the knowledge points distribution belongs to each subject and its course

Dataset	Relevance Quality		Image Quality			Text Quality		
	Img-Text Relevance	Img-Img Similarity	Unique Img%↑	Aesthetic Score↑	NSFW Score↓	Adv. Score↓	NSFW Score↓	Fluency (PPL) ↓
MMC4	0.53	0.31	60%	4.9	0.20	0.18	0.17	14.4
OBELICS	0.65	0.34	83%	**6.6**	0.24	0.15	0.05	**11.2**
Ours	**0.79**	**0.68**	**86%**	3.4	**0.18**	**0.11**	**0.02**	12.7

Table 8. Comparison of datasets on 8 metrics.

concepts of velocity and acceleration in physics. Figure 8 demonstrates the step-by-step, frame-by-frame problem-solving process for a mathematical geometry problem, detailing each critical step with accompanying text and visuals. Figure 9 presents a detailed depiction of chemical concepts such as atoms, molecules, and compounds through a combination of text and illustrations. Figure 10 introduces the depth-first search algorithm using an animation.

Except for refined ASR texts, we also provide the OCR texts in our textbook, which can be helpful for math-related scenario. For example, in Fig. 7, we utilize OCR to recognize formulas and symbols displayed on the screen, which facilitates better comprehension of physical concepts.

11. Ethical discussion

During the collection and release of our multimodal textbook dataset, We are very concerned about ethical considerations. In addition to following the established corpora (e.g., MMC4 [61], OBELICS [23] and Omnicorpus [26]), we make additional efforts to uphold high ethical standards, such as employing LLMs to filter out inappropriate videos, including those with biases, pornographic content, or personal privacy information, such as identification documents and bank account details. We are open to further refining our strategy while maintaining open-source resources based on community feedback.

12. License and Author Statement

We release the dataset under a CC-BY license and Terms of Use that require disclosure of when the dataset is used for the purpose of training models. This license is not intended to replace the licenses of the source content, and any use of content included in the dataset must comply with the original licenses and applicable rights of its data subjects.

The purpose of this statement is to clarify the responsibilities and liabilities associated with the use of this dataset. While we have made every effort to ensure the accuracy and legality of the data contained within this dataset, we cannot guarantee its absolute completeness or correctness.

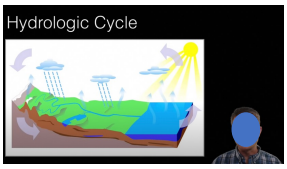
Therefore, if any rights, legal or otherwise, are violated through this dataset, including but not limited to copyright infringement, privacy violations, or misuse of sensitive information, we, the authors, assume no liability for such violations.

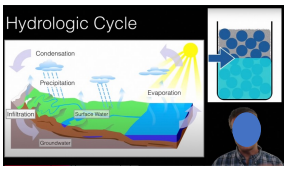
By utilizing this dataset, you agree that any consequences, legal or otherwise, arising from using this dataset will be the user's sole responsibility. You acknowledge that you will exercise due diligence and adhere to all applicable laws, regulations, and ethical guidelines when using the dataset.

By accessing, downloading, or using this dataset, you signify your acceptance of this statement and your commitment to abide by the terms and conditions of the CC-BY license.

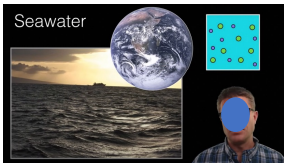
If you disagree with the terms of this statement or the CC-BY license, you are not authorized to use this dataset.The dataset will be hosted and maintained on the Hugging Face Hub.

ASR The hydrologic cycle works like this: wherever there is water on the surface, evaporation can occur. This water vapor eventually cools, leading to condensation and precipitation. Once the water reaches the surface, we call it surface water. This includes water running over the surface in lakes, swamps, and rivers. However, when water hits the ground, it can also infiltrate into the soil and underlying ground.

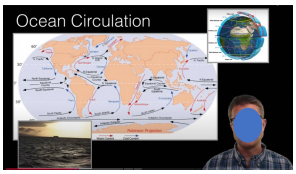




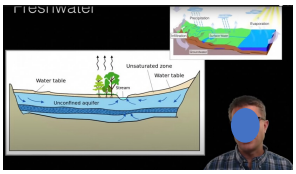
ASR To illustrate infiltration, imagine a beaker. Is the beaker full? It might appear full of air, but if you add marbles, there's still space between them. If you then fill it with sand, you'll see even smaller spaces remain. Finally, if you pour water into the beaker, it begins to truly fill up. This is what infiltration looks like—water flowing down into the soil, eventually becoming groundwater. The point at which the soil or rock becomes fully saturated is called the water table.



ASR Most of the Earth's surface is covered in water, but unfortunately, the majority is seawater, which contains dissolved salt. Drinking seawater is lethal, and it's unsuitable for crops. Ocean water circulates through currents, which are influenced by atmospheric patterns, Coriolis effects, and differences in salinity. For example, areas with high evaporation rates have higher salt concentrations, while regions with melting glacial ice have lower salt concentrations. These variations in salinity, combined with temperature differences, drive thermohaline circulation, connecting the entire ocean system into one cohesive flow.



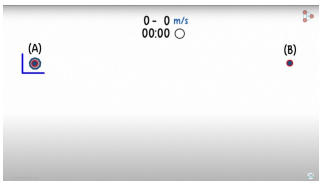
ASR When it comes to freshwater, it can be categorized as either surface water or groundwater. Surface water is above the ground, like streams and lakes. Groundwater, on the other hand, lies beneath the surface. You can see evidence of the groundwater table near streams—if you dig a hole beside the stream, it will fill with water due to the surrounding groundwater.

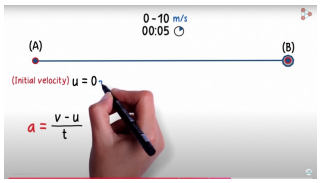


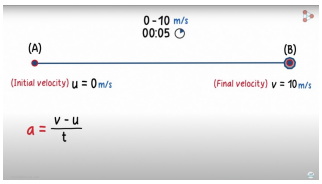
ASR This brings us to aquifers, which are underground storage areas for groundwater. An unconfined aquifer allows water to move freely between the surface and the aquifer. If you dig deeper, you might encounter a confined aquifer, which is trapped between impermeable layers of rock, restricting its movement. This distinction is crucial for understanding how water is stored and accessed beneath the ground.

Figure 5. A case presented in our textbook illustrates the water cycle within the domain of earth science.

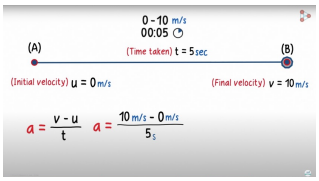
ASR: Assume the object is initially at rest at point A. It then moves to the right, reaching a velocity of 10 meters per second in 5 seconds. What would be the acceleration at point B?



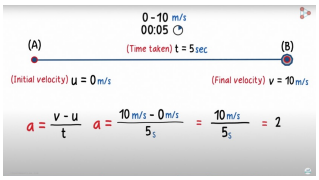


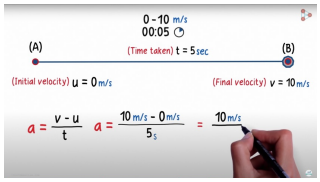


ASR: To find the answer, we can refer to the formula. The initial velocity, denoted as u, is 0 meters per second since the object is stationary. The final velocity, v is 10 meters per second, and the time taken, t is 5 seconds.

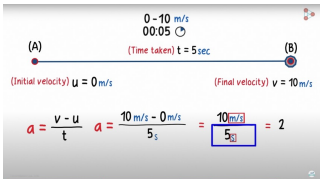


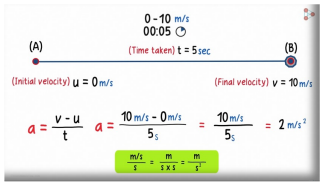
ASR: Therefore, the acceleration, A can be calculated by subtracting the initial velocity from the final velocity and dividing the result by the time taken. This gives 10 meters per second minus 0 meters per second, divided by 5 seconds. Performing this calculation, we have 10 divided by 5, which equals 2.





ASR: Now, let's determine the units of acceleration. Since acceleration is defined as the change in velocity divided by the time taken, its units will be meters per second per second. This can be simplified by expressing it as meters per second squared, which is more concise. Thus, in this case, the acceleration is 2 meters per second squared.



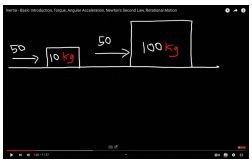


OCR

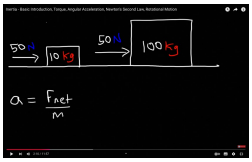
0 - 10 m/s 00:05 (Time taken) t = 5 sec
(A) (Initial velocity) u = 0 m/s
a = (v - u) / t
a = (10 m/s - 0 m/s) / 5 s
= 10 m/s / 5 s
= 2 m/s^2
(B) (Final velocity) v = 10 m/s
m/s/s = m / s x s = m / s^2

Figure 6. A case presented in our textbook introducing the principles of mechanics within the domain of physics.

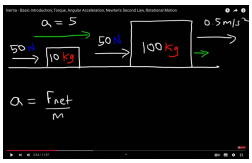
ASR To illustrate the concept of inertia, let's consider two objects. The first object has a mass of 10 kilograms, while the second object has a mass of 100 kilograms. In both cases, we apply a force of 50 newtons.



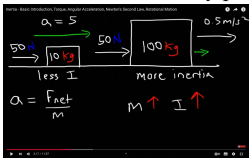
ASR Now, which object do you think has more inertia—the one with less mass or the one with more mass? Intuitively, you know it's the object with more mass. It's easier to move a lighter object, but more difficult to move a heavier object because the heavier object has greater inertia. This demonstrates that inertia is directly proportional to mass.



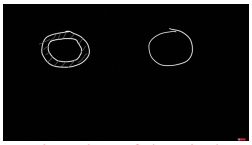
ASR Now, according to Newton's second law, the net force acting on an object is equal to the product of its mass and acceleration. From this, we can see that acceleration is the net force divided by the mass. For the first object, with a force of 50 newtons and a mass of 10 kilograms, the acceleration is 50÷10, which equals 5 meters per second squared.
For the second object, with the same force of 50 newtons but a mass of 100 kilograms, the acceleration is 50÷100, which equals 0.5 meters per second squared.



ASR As you can see, the lighter object has a much larger acceleration, while the heavier object has a much smaller acceleration. This indicates that the lighter object has less inertia, as it was easier to accelerate with a small force. In contrast, the heavier object has more inertia, since applying the same force resulted in a much smaller acceleration. Therefore, as the mass of an object increases, its inertia also increases. This demonstrates that inertia is directly proportional to mass.



ASR That's the basic concept of inertia as it applies to translational motion. But what about rotational motion? How does inertia come into play there? Let's compare two objects to explore this: the first is a thin hoop, and the second is a solid disk.



ASR For the thin hoop, the mass is concentrated along the edge of the circle, whereas for the solid disk, the mass is distributed throughout the circle. Now, let's assume that both objects have the same mass—10 kilograms—and the same radius of 2 meters. With these conditions, which one do you think has more rotational inertia? Is it the thin hoop or the solid disk? What would be your answer?

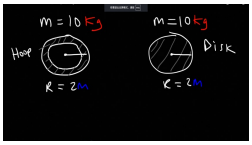
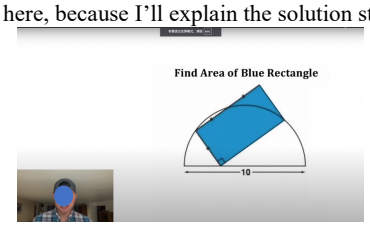
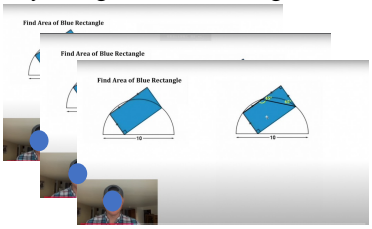


Figure 7. A case presented in our textbook introducing the concepts of velocity and acceleration within the context of physics.

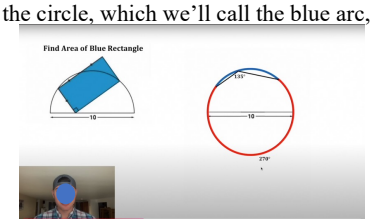
ASR Let's find the area of the blue rectangle. The rectangle is twice as wide as it is tall and intersects the semicircle at specific points. If you'd like to solve this on your own, pause here, because I'll explain the solution step by step.



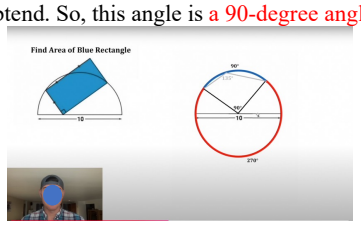
ASR First, let's connect two points on the rectangle and focus on the triangle that forms. This triangle is a right triangle because the shape is a rectangle, and it's also isosceles since two of its sides are congruent. This tells us it's a 45-45-90 triangle. Since one of its angles is 45 degrees, we can deduce another angle by noting that these two angles form a linear pair, meaning they add up to 180 degrees. That makes the other angle 135 degrees.



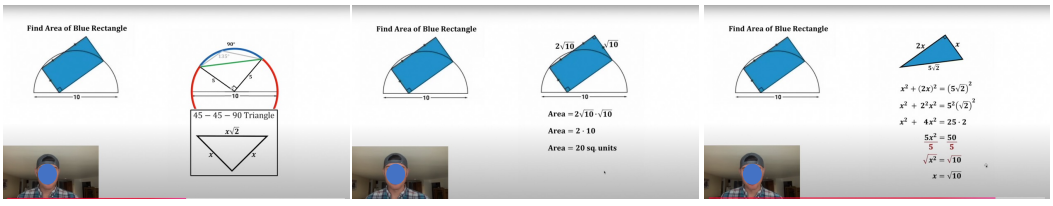
ASR Next, let's extend the circle and focus on the 135-degree angle. This angle is an inscribed angle, and inscribed angles have a special property—they are exactly half the measure of the arc they subtend. So, the red arc cut by this angle measures double 135 degrees, which is 270 degrees. The remaining arc of the circle, which we'll call the blue arc, must then be 90 degrees.



ASR Now, let's draw two radii to mark this 90-degree blue arc. The angle formed at the center of the circle is called a central angle, and central angles are equal to the arcs they subtend. So, this angle is a 90-degree angle.



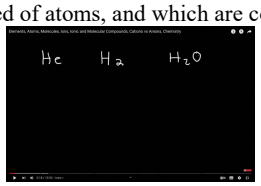
ASR Since both sides of this angle are radii, their lengths are equal to the radius of the circle, which is 5. Connecting the points forms another isosceles right triangle. Using the properties of a 45-45-90 triangle, we can determine that the hypotenuse of this triangle is 5\sqrt{2}. This value represents the diagonal of the rectangle. Returning to the rectangle, we know the shorter side is x, and the longer side, being twice as long, is 2x. Using the right triangle formed by these sides and the diagonal, we can determine the value of x. Solving this, we find that the shorter side of the rectangle is \sqrt{10}, and the longer side is 2\sqrt{10}.

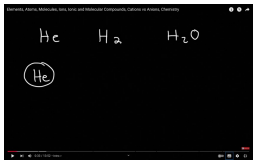


ASR Now, we can calculate the area of the rectangle. The area is equal to the base times the height. With the base as 2102\sqrt{10}210 and the height as \sqrt{10}, the area becomes 2×10, which is 20. Adding a label, the area of the blue rectangle is **20 square units**.

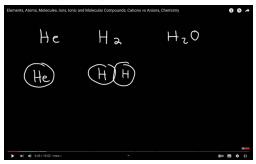
Figure 8. A case presented in our textbook demonstrates how to solve a question about planar geometry in the domain of mathematics.

ASR What is the difference between an atom and a molecule? Let's consider three substances: helium, hydrogen gas, and water (H_2O). Which of these are composed of atoms, and which are composed of molecules?

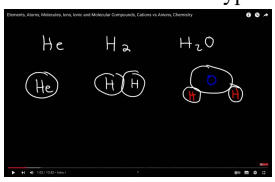




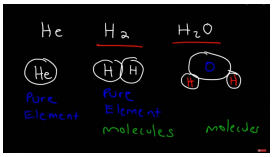
ASR Helium is composed of atoms. Each particle of helium consists of a single helium atom. In contrast, hydrogen gas is composed of molecules. Each particle of hydrogen gas contains two hydrogen atoms bonded together.



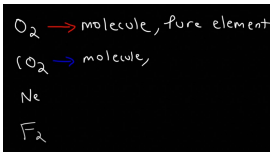
ASR A molecule, by definition, is a particle made up of two or more atoms. Water is also composed of molecules, specifically one oxygen atom and two hydrogen atoms. At the center of a water molecule is the larger oxygen atom, with the two smaller hydrogen atoms attached. Thus, a molecule can consist of either the same type of atom or different types of atoms bonded together.



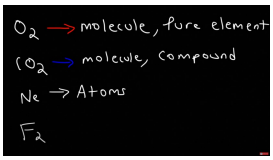
ASR Helium is classified as a pure element because it is made up of only one type of atom. Similarly, hydrogen gas (H_2) is a pure element, as it is composed solely of hydrogen atoms. Water, however, is not a pure element; it is a compound, as it contains two different types of atoms: hydrogen and oxygen. While H_2 and H_2O are both considered molecules, only H_2 is a pure element, and H_2O is classified as a compound due to the presence of multiple atom types.



ASR Now, let's work through more examples. For each substance, determine whether it is composed of atoms or molecules and whether it is a pure element or a compound.



ASR Let's start with O_2. Is it composed of atoms or molecules? A particle of O_2 consists of two oxygen atoms bonded together, so it's a molecule, as it contains multiple atoms. Since it only contains one type of atom—oxygen—it is classified as a pure element. Now, what about CO_2? Would you classify a particle of CO_2 as an atom or a molecule? Since CO_2 is made up of multiple atoms, it is considered a molecule.

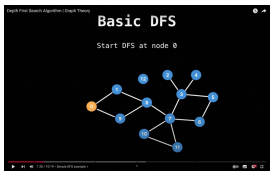


ASR However, it doesn't consist of just one type of atom. It contains both carbon and oxygen atoms, meaning it is not a pure element. Instead, it is a compound. A compound is defined as a substance composed of different types of atoms bonded together. In other words, a compound is not a pure element.

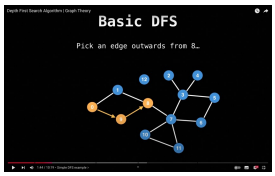
Finally, let's look at neon. Neon is composed of individual atoms. A particle of neon consists of a single neon atom, unlike oxygen molecules, which are made up of two oxygen atoms bonded together. Neon, therefore, is not a molecule but is instead classified as an atom and a pure element, as it contains only one type of atom.

Figure 9. A case presented in our textbook illustrates the concepts of molecules, atoms, and compounds in the domain of chemistry.

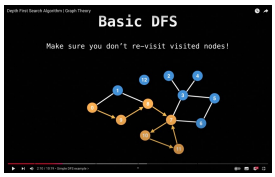
ASR Depth-first search (DFS) works by selecting the next node to explore until it cannot proceed further, at which point it backtracks and continues its exploration from a previous point. Let's start a depth-first search on node 0 and see how it unfolds. We begin at node 0 and arbitrarily choose a node to move to.



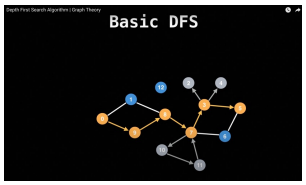
ASR From node 0, we go to node 9. At node 9, there's only one option, so we move to node 8.

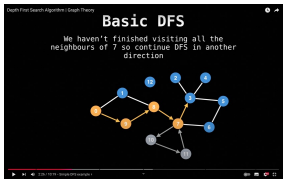


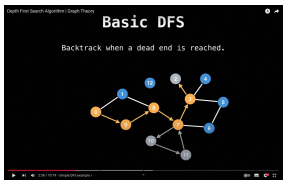
ASR At node 8, we arbitrarily select an edge and proceed to node 7. At node 7, there are multiple edges to choose from, so we decide to go to node 10. From node 10, we move to node 11, and then back to node 7. However, we do not revisit already visited nodes or nodes that are currently being explored, so at this point, we need to backtrack. To indicate backtracking, we'll label the edges and nodes as gray. We backtrack all the way to node 7. Since we haven't finished exploring all edges from node 7, we pick another edge and move to node 3.



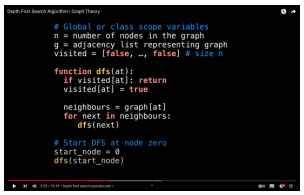
ASR At node 3, we proceed to node 2. Node 2 is a dead end, so we backtrack and explore another edge from node 3, moving to node 4. Node 4 is also a dead end, so we backtrack again to node 3. Finally, we take the last edge from node 3 and move to node 5. From node 5, we go to node 6, and from node 6, we reach node 7 again. Since node 7 is currently being visited, we backtrack all the way to node 8. At node 8, there's still one unvisited edge that leads to node 1. From node 1, we move back to node 0. However, since node 0 is currently being explored, we backtrack all the way to node 0. At this point, we've completed the depth-first search traversal of the graph.







ASR This was just one possible depth-first search traversal, as the path can vary depending on the choices made during exploration. Now, let's look at some pseudocode for depth-first search to gain a deeper understanding of its implementation. The first step we'll need is to...



OCR "# Global or class scope variables\nn = number of nodes in the graph\ng = adjacency list representing graph\nvisited = [false, …, false] # size n\n\nfunction dfs(at):\n if visited[at]: return\n visited[at] = true\n\n neighbours = graph[at]\n for next in neighbours:\n dfs(next)\n\n# Start DFS at node zero\nstart_node = 0\ndfs(start_node)"

Figure 10. A case presented in our textbook introduces a depth-first search algorithm.