

A. Data Annotation

Number of images	39,153
Number of human annotations	39,153
Average # of words per human annotations	65.1
Number of QA pairs	102,437
Average # of words per question	10.5
Average # of words per answer	23.9

Table 7. TRINS datasets summary

B. TRINS-Gen: Text-to-Image Generation

Text-Rich Image Generation Diffusion-based Text-to-image Generation has shown great success, while precise textual renderings remain a big challenge. Imagen [42], eDiff-I [4], and DeepFloyd [44] exploit large T5 models [40] for better precision. The study by Liu et al. [28] proposes a character-sensitive alternative to traditional token length handling. GlyphDraw [30] merges superior images with Chinese texts, while Textdiffuser [6] uses the Transformer model for multiline text generation and segmentation masks for user control. However, manual input is still needed for key-term recognition. TextDiffuser also introduced the MARIO-Eval benchmark, drawing from works such as DrawBench [42] and DrawTextCreative [28]. However, most text prompts in MARIO-Eval are short and cannot serve as a good evaluation dataset to handle complex real-world human instructions. In TRINS-Gen, we introduce a new human-annotated image generation benchmark for existing text-rendering methods, together with a new training dataset.

Text rendering is a main challenge for text-rich image generation, and we do an extensive evaluation of existing methods on the TRINS-Gen benchmark. We take advantage of human annotations and build the TRINS-Gen benchmark. We have chosen data examples with fewer words, as existing methods cannot yet render too many texts within a single image. Following Chen et al. [6], we consider the calculated FID score between the generated images and the TRINS-Gen ground truth images. In addition, we calculate the CLIP scores to measure whether the visual components within the generated image followed the text instructions, and OCR accuracy to measure whether the desired texts are generated correctly. We evaluated existing methods using their public checkpoints and reported the results in Table ???. We can see that DeepFloyd has the best image quality and aligns well with the user-text prompts. ControlNet, which takes a rendered black-white text image as a condition, has an image quality similar to that of TextDiffuser. In terms of text rendering ability, TextDiffuser is best compared to other baselines. In Figure 15, we show two examples of images generated by different methods. Stable Diffusion (SD) models cannot render any meaningful text. Deepfloyd and ControlNet may miss some words or render texts with spelling errors. TextDiffuser can generally render text well

when the length of words is short. However, all methods do not render the desired texts when the number of words increases, as described in the Appendix C.3. The main cause in TextDiffuser [6] is that its layout models cannot parse complex prompts, and an LLM-based layout model may help.

B.1. DocGen: Text-rich Image Generation

Given high-quality text-rich image descriptions, it is natural to consider their application in training and evaluating text-rich image generation models [6]. As highlighted in Section 3.3, the average number of OCR words is 65.1. However, generating such a substantial amount of text within a single image poses a significant challenge for diffusion models. In response to this, we have established the first human-annotated text-rich image generation benchmark. This benchmark involves filtering out images with more than 20 OCR words, resulting in a curated set of 2,104 images. We divide these images into two sets based on the number of OCR words and the length of the longest OCR string per annotation, where all text prompts in the easy set have less than 9 OCR words with the longest OCR string of 5. Visual examples of both data sets are included in Appendix A.

C. Additional Experiments

C.1. More TRINS-Cap Results

Figure 16 shows more results of random examples from TRINS-Cap.

C.2. More TRINS-QA Results

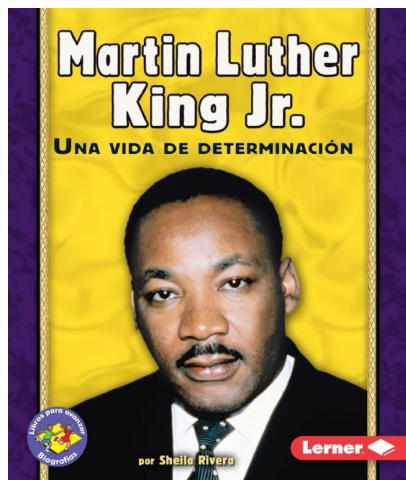
Figure 18 shows more random examples of the extract QA task and Figure 17 shows two more random examples of the abstract QA task.

C.3. More TRINS-Gen Results

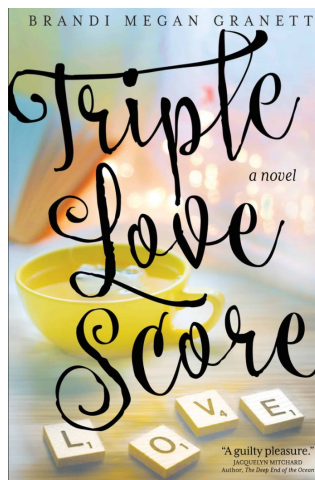
Figure 9 shows the results of different methods on the TRINS-Gen difficult benchmark. Deeploy model needs more than two A100-80GB GPUs and we omit its result due to the resource limitation. SD-2.0 model demonstrated the best performance in image quality and alignment with text descriptions, outperforming SD-1.5, ControlNet, and TextDiffuser. TextDiffuser, however, showed its strength in accurately generating text within images, leading in metrics related to text recognition. ControlNet exhibited balanced performance in these text-related tasks, making it a notable contender in terms of precision and recall balance. In general, when text prompts become complex, SD-2.0 and SD-1.5 can still generate meaningful images related to the prompts. Textdiffuser struggle to place to many words in the image and fail to generate good images. In addition, its text generation quality is much lower as the number of generated words increases. Based on the results, it is obvious to see



Figure 10. TRINS datasets annotation statistics



The image is of a book cover for a biography of Martin Luther King Jr. written in Spanish. The book cover is predominantly yellow with a purple border. The title of the book is "Martin Luther King Jr. Una vida de determinación" written in white and black text. The author's name, "por Sheila Rivera", is written in smaller yellow text at the bottom of the cover. The cover features a photo of Martin Luther King Jr. The bottom right corner of the cover has a small logo for "Lerner" publishing.



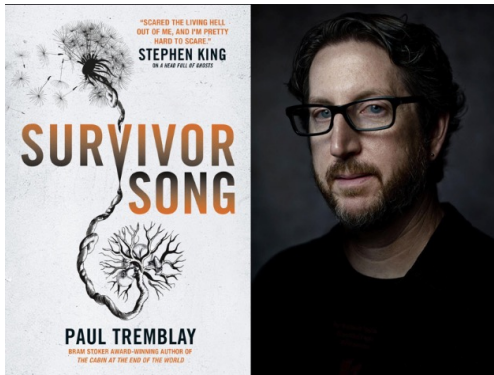
This is the cover of the book titled "Triple Love Score." The title is centrally positioned, written in a cursive black font with a prominent size, leaning towards big. Right above the title, the author's name "Brandi Megan Granett" is presented in a smaller uppercase font. Directly below the title, the phrase "a novel" is subtly written, indicating the genre of the book. At the bottom, there's a quote that reads, "A guilty pleasure." attributed to Jacquelyn Mitchard, with an additional mention: "Author, The Deep End of the Ocean." The most captivating aspect of the background is an inviting and warm scene. It portrays a bright yellow bowl filled with liquid, with the word "LOVE" spelled out using tile letters on a table.

Figure 11. TRINS-Cap Human Annotated Examples

that TRINS-Gen difficult evaluation set is more challenging and need to be better tackled by future methods.

D. More Annotation Examples

Figure 10 shows the annotation statistics from the LabelBox.



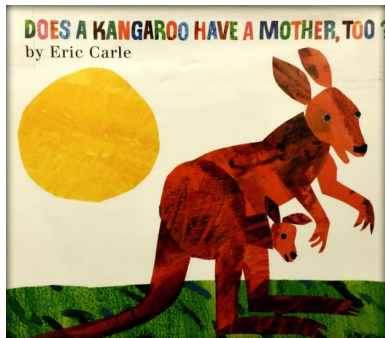
This is a book cover. The book cover is pictured next to the portrait of the author who is a man with a beard wearing a black shirt and black glasses. Text in the upper right of the cover reads "scared the living hell out of me, and I'm pretty hard to scare. Stephen King on a head full of ghosts". The title " survivor song" is in large orange and black radiant text in the center of the cover with subtext below in the bottom center that reads "Paul tremblay Bram Stoker award-winning author of the cabin at the end of the world". The cover image is a black line illustration of a neuron turning into a dandelion with seeds blowing away in the top left.



This is an image of four DVD plates with the title "LAND OF PROMISE: THE BRITISH DOCUMENTARY MOVEMENT 1930-1950". At the left side of the plate is an acronym "BFI", at the right side of the plate is written, "Disk 4". On the plate is a picture of a man carrying stick on his shoulder. The plates is flagged at the left and right sides with the cover of the plates. On the left cover is photo of people sitting and looking up,. On the right cover is an image of two children sitting and reading. The texts on the plate is the replica of that on the covers. The background colour of the plates and covers is black.



This image appears to be a visual representation discussing the relationship between humans and horses. The title "Do horses like humans?" stands out prominently, positioned at the bottom. The font color is black, and the size of the title is big. Below the title, there's an additional line that says, "Studies show they understand our emotions," in a smaller font, also in black and located at the very bottom of the image. The prime focus in the background showcases a close-up of a young girl giving a gentle kiss to a horse's snout

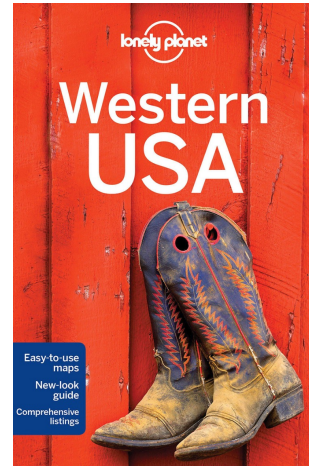


This is a book cover. The title "does kangaroo have a mother too?" is enlarged colorful text in the top with subtext under it that reads "by eric Carle". The cover image is a mother kangaroo with a Joey in her pouch standing on green grass and behind a yellow sun done in a watercolor paper cut style.

Figure 12. TRINS-Cap Human Annotated Examples



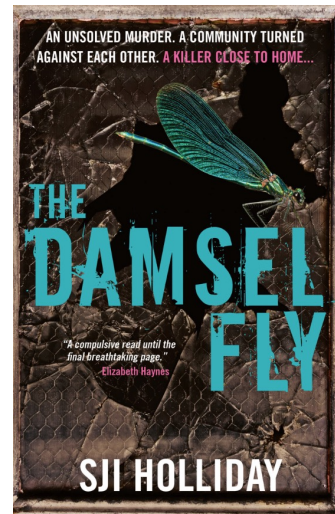
This is a book cover. Text in the upper left reads "A Golden Book". The title "The Big Golden Book of Backyard Birds" is in large blue and red text on the upper center with subtext underneath that reads "text by Kathleen N. Daly". The cover image is an illustration of a variety of birds in a backyard scene with sparrows and white flowers in the foreground in the left and a cardinal sitting on top of a wooden post with ivy growing on it in the foreground on the lower right.



This is a book cover. Text at the top reads "Lonely Planet". The title "Western USA" is in large white text in the upper center. Blue rectangle in the bottom left contains subtext that reads "Easy to Use Maps New-look guide Comprehensive listings". The cover image is a photograph of brown and black and red cowboy boots hanging on a hook on a red distressed wooden wall.



The image is of a book cover for "The Adventures of Tintin and the Picaros" by Hergé. The cover is illustrated in a comic book style and features Tintin, his dog Snowy, and two other characters in a jungle setting. The background is a pyramid temple with red flowers and green foliage. The title of the book is in a yellow speech bubble at the top of the cover, and the publisher's logo is at the bottom of the cover.



The image is of a book cover for the novel "The Damsel Fly" by SJI Holliday. The cover is black with a large blue-green dragonfly on the right side. The title of the book is written in large blue letters, and the author's name is written in smaller white letters at the bottom of the cover. The top of the cover has a quote in white letters that reads "An unsolved murder. A community turned against each other. A killer close to home...". The cover has a distressed and torn look around the edges.

Figure 13. TRINS-Cap Human Annotated Examples



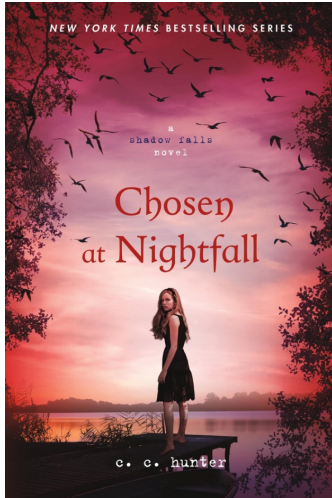
Question: What is the title of the book in the image?
Answer: The Big Golden Book of Backyard Birds.

Question: Who is the author of the book?
Answer: Kathleen N. Daly.



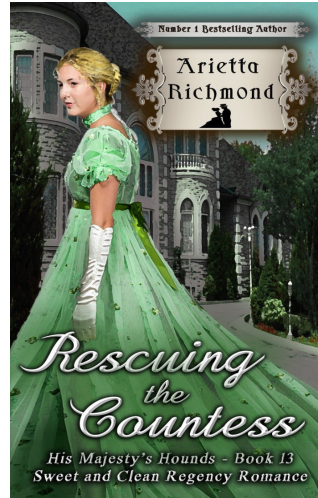
Question: Who is the author of the book shown in the image?
Answer: Tim Flannery.

Question: Who is the illustrator of the book?
Answer: Sam Caldwell.



Question: Who is the author of the book displayed in the image?
Answer: C.C. Hunter.

Question: What is the title of the book?
Answer: Chosen at Nightfall



Question: Who is the author of the book depicted in the image?
Answer: Arietta Richmond.

Question: What is the title of the book?
Answer: "Rescuing the Countess: His Majesty's Hounds - Book 13: Sweet and Clean Regency Romance".

Figure 14. TRINS-QA Human-assisted Annotated Examples

Text Prompts: The image is a book cover for 'Abandoned San Diego' by Jessica D. Johnson, featuring a background image of an abandoned house.



Text Prompts: The image is a book cover titled 'The Little Princess', featuring the image of a smiling little girl. The title and the image are presented in vibrant colors, suggesting a lively and joyful atmosphere.



Figure 15. Examples generated by different text-to-image models on the TRINS-Gen benchamrk.

Metrics	SD-2.0	SD-1.5	ControlNet	DeepFloyd	TextDiffuser
FID (↓)	59.54	59.05	51.59	49.96	51.26
CLIP Score (↑)	0.3520	0.3419	0.3717	0.3917	0.3707
OCR Accuracy (↑)	0.0893	0.2205	0.4241	0.2192	0.5027
OCR Precision (↑)	0.1019	0.2535	0.4083	0.2225	0.6001
OCR Recall (↑)	0.0900	0.2218	0.4268	0.2669	0.5061
OCR F-measure (↑)	0.0956	0.2366	0.4174	0.2427	0.5491

Table 8. Empirical Results of different methods on TRINS-Gen (easy) benchmark.



GT: The image is of an audio book cover for "Just One Kiss" by Susan Mallery. The cover is a photo-realistic image of a couple laying on a grassy field in a romantic embrace. The title of the book is written in a large blue font at the center right of the cover, and the author's name is in a larger green font at the top. The cover also includes a quote from #1 New York Times bestselling author Debbie Macomber "Susan Mallery is one of my favorites." and a tagline for the book "A fool's Gold Romance". There is also BrillianceAudio logo at the top left and Unabridged at the top right.

Instruct-BLIP: Just One Kiss is a romance novel written by Susan Mallery. In this image, the cover of the book is prominently displayed in the foreground, with the title "Just One Kiss" and the author's name, Susan Mallery, clearly visible. The couple on the cover is shown in an intimate embrace, with the woman leaning over the man's shoulder. There are several other books scattered throughout the image, but they are not as prominent as the cover of Just One Kiss.

LLaVAR: In the image, a couple is seen lying down in the grass, sharing an intimate moment with each other. They are positioned next to each other, and it seems like they are about to kiss. The scene is set with a backdrop of grass and flowers, making it a romantic and serene environment. Additionally, there are a few words scattered throughout the scene, likely related to the quote or description of the book. There is also a small cat situated at the top right corner of the image, possibly observing the couple or enjoying the outdoors.

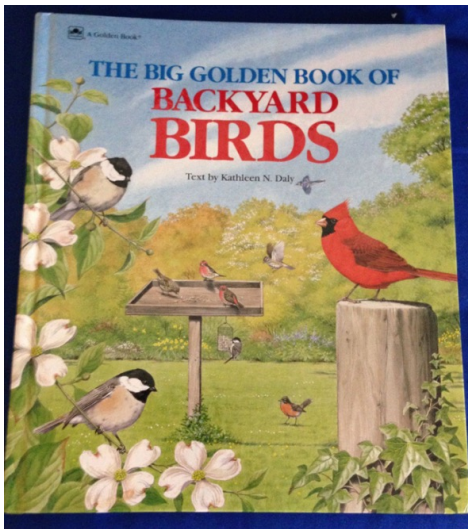
LLaVA: The image depicts a man and a woman sitting in a grassy field, enjoying a summer day. They are positioned close to each other, creating an intimate atmosphere. The woman has her head resting on the man's shoulder, and both seem to be relaxed in each other's company. A book, likely a paperback copy of Susan Mallery's "Just One Kiss," is visible in the scene, suggesting that they might be spending their day reading and bonding outdoors. There is also a clock in the background, indicating they are keeping track of the time during their picturesque summer day.

mPLUG-Owl2: The image features a man and a woman sitting on a beach. They are both wearing swimsuits, and the woman has a handbag beside her. The scene appears to be a romantic moment between the two individuals, enjoying their time together on the beach.

Qwen-VL: The image shows a couple lying on the grass, facing each other and kissing. The man is wearing a yellow shirt and is in the process of removing his sunglasses. The woman has long dark hair and is wearing a light blue dress. Her left hand is placed on the man's chest, while his right hand is holding hers. They both have light skin and are smiling.

LaRA w/o OCR (zero-shot): The image displays the cover of a romance novel called "Just One Kiss" by Susan Mallery. The book is a New York Times Bestseller, and the author is recognized as a #1 New York Times Bestselling Author. The text on the image is a quote from Debbie Macomber, another renowned author, who states that Susan Mallery is one of her favorite authors. The quote is attributed to the New York Times bestselling book "Just One Kiss". The cover also features an image of a couple engaged in an intimate moment.

LaRA: This is the cover of the audiobook named Just One Kiss: A Fool's Gold Romance. Here the title is written at the lower part with large fonts. At the upper part it mentions the author Susan Mallery. Just before that it says, New York Times Bestselling Author. Under the author's name it includes an endorsement. At the top right it includes a logo of Unabridged and at the top left it includes the logo of Brilliance audio. The cover includes an image of an intimate couple as the backdrop.



GT: This is a book cover. Text in the upper left reads "A Golden Book". The title "The Big Golden Book of Backyard Birds" is in large blue and red text on the upper center with subtext underneath that reads "text by Kathleen N. Daly". The cover image is an illustration of a variety of birds in a backyard scene with sparrows and white flowers in the foreground in the left and a cardinal sitting on top of a wooden post with Ivy growing on it in the foreground on the lower right.

mPLUG-Owl2: The image features a book titled "The Big Golden Book of Backyard Birds." The book is placed on a blue background, and there is a bird on a post in the foreground.

LLaVA: The image features a beautifully illustrated children's book about birds. The book cover depicts a wooden birdhouse hanging on a tree, surrounded by various birds perched around it. Some birds are sitting on the birdhouse, while others are scattered across the scene, showing the lively nature of the illustration. The book itself likely contains stories or information about these birds and their habitats. The illustration is bright and colorful, capturing the essence of nature and attracting the attention of young readers.

LLaVAR: The image features a wooden birdhouse mounted on a wooden post. A bird can be seen perched on the bird feeder, making it an inviting sight for birds to visit. There are a few books in the scene, with one large book taking up much of the space on the right side, and another smaller book positioned on the left side. The books seem to be related to birds or nature, possibly providing information about the birds that might visit the birdhouse.

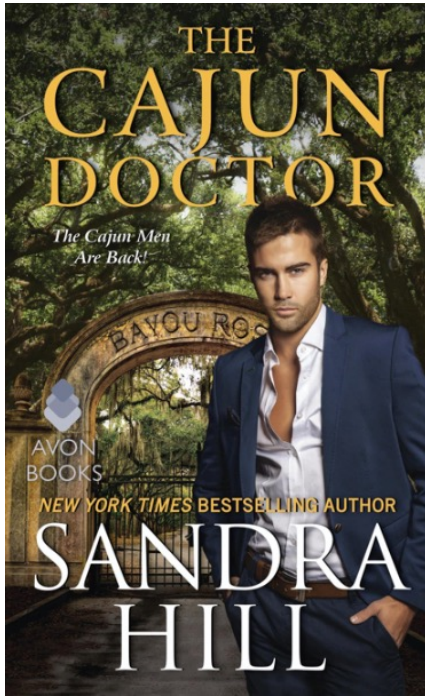
Instruct-BLIP: A book titled "The Big Golden Book of Backyard Birds" is displayed on a blue background. The book showcases various types of birds, including cardinals, robins, blue jays, and sparrows, among others.

Qwen-VL: The image features a blue book cover with the title "The Big Golden Book of Backyard Birds" written in red and gold letters. There are several birds mentioned in the book, as illustrated by the various birds featured on the cover. A cardinal is perched on a wooden post, while a black-capped chickadee is at the left side of the cover. A blue jay can be seen flying above a birdhouse, and a sparrow is standing on the ground near the bottom of the cover. The book is placed on a blue surface with green plants on the right side of the cover, adding a touch of nature to the scene.

LaRA w/o OCR (zero-shot): The image features the cover of a book titled "The Big Golden Book of Backyard Birds". The book has a blue cover with a white border. The title is written in a mix of blue and red text. There is a bluebird illustration on the cover, and it is sitting on a birdhouse. The birdhouse is mounted on top of a wooden post. The background of the cover shows a backyard scene with trees and grass.

LaRA: This is a cover of a book titled "THE BIG GOLDEN BOOK OF BACKYARD BIRDS" The text on the cover is written in a light blue color, with a black outline. At the center of the cover is written "Text by Kathleen M. Daly" At the top of the cover is a logo of a tree with branches and goldens. "A Golden Book" is written in small letters beneath it The book cover has a picture of a bird house with birds flying in and out in the background. The background color of the book cover is green. The book is placed on a blue background.

Figure 16. More Generated Examples by different methods on TRINS-Cap.



Question: What is the title of the book shown in the image?

Answer: The Cajun Doctor.

InstructBLIP: The Cajun Doctor

LLaVAR: The title of the book is "The Cajun Doctor" by Sandra Hill.

mPLUG-Owl2: The title of the book is "The Cajun Doctor" by Sandra Hill.

Qwen-VL: The title of the book shown in the image is "The Cajun Doctor".

LaRA: The title of the book is "The Cajun Doctor".

Question: What is the name of the publisher as shown on the book cover?

Answer: Avon Books.

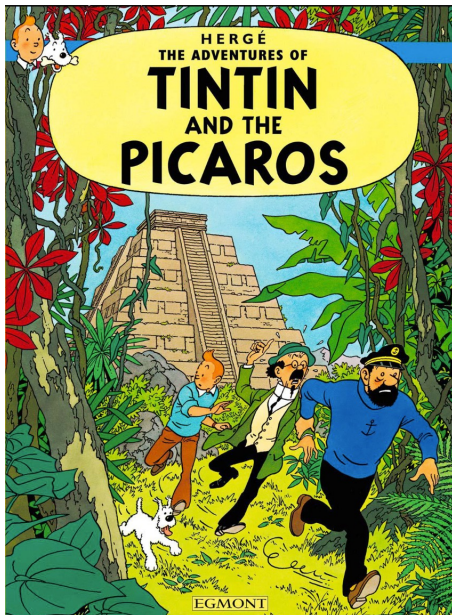
InstructBLIP: St. Martin's Paperbacks

LLaVAR: The publisher is Avon Books, a division of HarperCollins Publishers.

mPLUG-Owl2: The publisher is Avon Books.

Qwen-VL: Avon Books

LaRA: The publisher's name is Avon Books.



Question: Who is the author of the book shown in the image?

Answer: Hergé.

InstructBLIP: Hergé

LLaVAR: The author of the book is Hergé.

mPLUG-Owl2: The author of the book is Hergé.

Qwen-VL: The author of the book shown in the image is Hergé.

LaRA: The author of the book is Hergé.

Question: What is the title of the book?

Answer: The Adventures of Tintin and the Picaros

InstructBLIP: The Adventures of Tintin and the Picaros

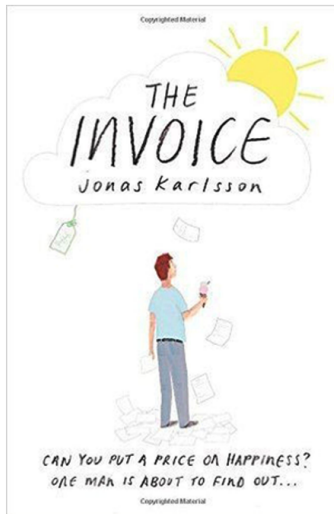
LLaVAR: The title of the book is "The Adventures of Tintin and the Picaros" by Hergé.

mPLUG-Owl2: The title of the book is "The Adventures of Tintin and the Picaros."

Qwen: The title of the book is "The Adventures of Tintin and the Picaros".

LaRA: The title of the book is "The Adventures of Tintin and the Picaros."

Figure 17. TRINS-QA Extract QA Results



Question: Can you tell me more about the illustration on the book cover?

GT: The illustration on the book cover depicts a man with an ice cream, looking at the clouds with the sun as the backdrop. This could be symbolic of the theme of the book, suggesting a search for happiness or a contemplation of its value.

Qwen-VL-Chat: The book cover features an illustration of a man standing under a cloud. He is holding a piece of paper in his right hand, which appears to be a bill or an invoice. There is a paper plane falling towards him, indicating that he might be receiving more documents. The man is wearing a light blue shirt and light-colored pants. He has short brown hair and is holding a glass in his left hand. The background is white, and there is a bright yellow sun on the top right corner.

mPLUG-Owl2: The book cover features a cartoon of a man and a cloud. This illustration likely represents the main character of the story, the man, who is in a contemplative or introspective state, possibly reflecting on the themes of happiness and the value of time. The cloud might symbolize the man's thoughts or ideas, as clouds often represent the unseen or abstract.

GPT-4V: The illustration on the book cover shows a stylized image of a man standing and looking at an invoice or paper in his hand. He appears to be in a thoughtful or puzzled pose, possibly contemplating the contents of the paper. Above him is a large, simple drawing of the sun, suggesting it might be daytime. There are also other papers, possibly more invoices or documents, floating around him, which may represent the overwhelming nature of the situation he is facing. The art style is cartoonish and minimalist, using a limited color palette and simple shapes to convey the scene. The overall impression is light and whimsical, despite the potentially serious subject matter suggested by the title "The Invoice." The tagline "Can you put a price on happiness? One man is about to find out..." hints at a narrative that explores themes of value, happiness, and perhaps the unexpected consequences of life's experiences.

LaRA: The book cover features an illustration of a man with ice cream, which might signify happiness or pleasure. The title "THE INVOICE" and the author's name, Jonas Karlsson, are prominently displayed. There's also a phrase "CAN YOU PUT A PRICE ON HAPPINESS? ONE MAN IS ABOUT TO FIND OUT," suggesting the theme or storyline of the book.

Question: Can you describe the design and color scheme of the cover image?

GT: The cover image has an abstract design featuring three silhouettes in green and blue colors. They are standing amidst what appears to be an abstract representation of a jungle, depicted with blue and black lines. The overall color scheme is predominantly green, blue, and black, creating a cool and mysterious atmosphere.

Qwen-VL-Chat: The cover image features a deep blue-green color scheme with large areas of blue and green brushstrokes. The background presents an abstract underwater scenery, with the title "ARJUNA OAKES" presented in orange at the top. The name of the EP, "The Watcher", is written in small font at the bottom. The artwork also includes small orange dots and red berries scattered throughout the painting.

mPLUG-Owl2: The cover image features a blue and green background with a bird on it. The bird is the main focus of the image, standing out against the vibrant colors of the background.

GPT-4V: The cover image presents a vivid and somewhat mystical scene. The primary color palette consists of various shades of blue and green, creating a nocturnal jungle atmosphere. In the foreground, there are ghost-like figures with a slightly glowing effect; these figures are in a lighter green hue, making them stand out against the darker background. Their eyes are represented by small white spots, and they seem to be peering out from the dense foliage. The jungle itself is represented by a multitude of leaves and plant forms, painted in deep blues and greens, which adds to the depth and complexity of the scene. There are also pops of orange, likely representing flowers or fruits, which provide a striking contrast to the cooler tones of the background. The text "ARJUNA OAKES" is prominently displayed at the top in a serif font in a light color, probably white or a very light yellow, ensuring legibility against the darker background. Below, in smaller lettering, is "THE WATCHER EP", which follows the same color scheme as the artist's name. The overall design conveys a feeling of watching and being watched in a mystical, natural setting, accentuated by the shadowy figures and the hidden eyes within the foliage. The use of contrasting colors and the glow effect adds to the ethereal quality of the image.

LaRA: The cover image for "The Watcher" by Arjuna Oakes features a mysterious and surreal scene. It depicts three ghostly figures in the woods, which creates an eerie atmosphere. The background predominantly consists of black and blue colors, further enhancing the mysterious and ominous vibe of the image.

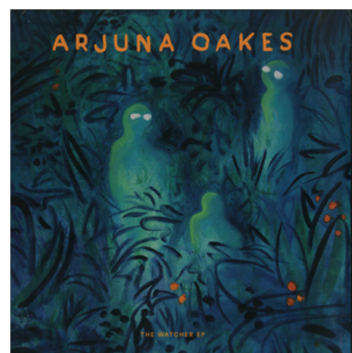


Figure 18. TRINS-QA Abstract QA Results

Metrics	SD-2.0	SD-1.5	ControlNet	TextDiffuser
FID (↓)	55.48	56.73	60.15	74.56
CLIP Score (↑)	0.3512	0.3384	0.3301	0.3219
OCR Accuracy (↑)	0.0447	0.0256	0.0833	0.0952
OCR Precision (↑)	0.0877	0.0513	0.1011	0.1249
OCR Recall (↑)	0.0450	0.0252	0.0827	0.0941
OCR F-measure (↑)	0.595	0.0338	0.910	0.1073

Table 9. Empirical Results of different methods on TRINS-Gen difficult benchmark.

	ST-VQA	OCR-VQA	TextVQA	DocVQA	ChartQA	InfoVQA	FUNSD	SROIE
BLIP-2 [22] †	21.7	30.7	32.2	4.9	3.4	11.3	0.20	0.14
OpenFlamingo [2] †	19.3	27.8	29.1	5.1	9.1	15.0	0.85	0.12
MiniGPT4 [62] †	14.0	11.5	18.7	3.0	4.3	13.3	1.19	0.04
mPLUG-Owl [55] †	29.3	28.6	40.3	6.9	9.5	16.5	1.02	0.60
LLaVA [27] †	28.9	11.0	36.7	6.9	28.9	13.8	1.02	0.12
LLaVA1.5 [27] †	38.1	58.1	38.7	8.5	9.3	14.7	0.20	1.70
LLaVAR †	39.2	23.8	48.5	11.6	12.2	16.5	0.50	5.20
mPLUG-Owl2 [56] †	29.3	28.6	40.3	6.9	19.4	18.9	1.40	3.20
LLaVAR w/ OCR	44.8	42.7	60.4	48.3	-	-	-	-
LLaVAR (finetuned)	40.3	28.8	50.1	12.2	-	-	-	-
LaRA	47.2	41.2	59.9	50.8	25.6	28.4	23.2	36.6

Table 10. Zero-shot performance (accuracy %) on text-based VQA. We use † to refer to the results obtained from Liu et al. [29].

Method	Resolution	Extract Accuracy	Abstract						
			B@1	B@2	B@3	B@4	METEOR	ROUGE	CIDEr
LLaVA [27]		22.6	12.3	7.5	4.9	3.4	15.3	20.2	26.2
LLaVAR [61]	336 ²	44.1	27.5	19.5	14.4	10.9	21.6	33.0	94.7
LLaVAR (finetuned)		50.1	26.1	18.7	14.1	11.0	23.1	33.4	112.6
LaRA		58.8	31.0	23.6	18.6	15.1	26.5	38.0	135.6

Table 11. Results of different models on TRINS-VQA-Human for text-rich image question-answering tasks.