

Technical Appendices and Supplementary Material

A NaViL-9B: Scaling up to 9B parameters

To further demonstrate the scaling capability of our method, we trained NaViL-9B based on Qwen3-8B [20]. The total number of activation parameters is 9.2B, of which 1.2B belongs to the visual encoder. The training recipe is similar to NaViL-2B, as shown in Tab. 4, except the visual multi-scaling packing is disabled in the first sub-stage of pre-training for acceleration.

Tab. 1 presents a comparison of the total training tokens required by our method versus two compositional counterparts. Notably, our approach achieves comparable performance while using substantially fewer training tokens, demonstrating improved training efficiency.

Table 1: Comparison between NaViL and existing MLLMs on the number of training tokens.

Models	Train ViT	Train MLLM	Total
Qwen2.5VL [1]	unknown	4.1T	>4.1T
InternVL2.5-8B [7]	>3.3T	140B	>3.5T
NaViL-2B (ours)	0	800B	800B
NaViL-9B (ours)	0	450B ¹	450B

The performance results on multimodal and visual question answering benchmarks are shown in Tab. 2. With a similar parameter size, our NaViL-9B outperforms all existing native MLLMs by a large margin on almost all benchmarks. Besides that, compared to the compositional baseline model InternVL-2.5-8B with a similar parameter size, NaViL-9B also achieves competitive performance. Such results show that our proposed native MLLM can be scaled up to larger parameter sizes and achieve consistent performance gains.

B Implementation Details

The detailed training hyper-parameters for NaViL-2B and NaViL-9B are provided in Tab 3 and Tab. 4, respectively. The high-quality multimodal data used in Pre-training and Supervised Fine-tuning is from InternVL-2.5 [7], which is sourced from various domains, such as image captioning, general question answering, multi-turn dialogue, charts, OCR, documents, and knowledge, *etc.*; while the pure language data is primarily from InternLM2.5 [4].

C Architecture of Modality-specific MoE

To facilitate visual pre-training while preserving pre-trained LLM knowledge, Modality-specific Mixture-of-Experts (MMoE) was introduced in Mono-InternVL [17] for building native MLLM. However, we empirically found that using only the feed-forward network (FFN) expert would lead to a significant difference in feature scale between visual and language modalities. To mitigate this issue, we further introduced modality-specific attention expert, that is, using different projection layers (*i.e.* qkvo) in the self-attention layer to process visual and text features respectively, and then perform unified global attention calculation. Specifically, the output $x_{i,m}^l \in \mathbb{R}^d$ of the i -th token with modality $m \in \{\text{visual, linguistic}\}$ at the l -th layer of the MoE-extended LLM can be defined as

$$\begin{aligned} x_{i,m}^{l'} &= x_{i,m}^{l-1} + \text{MHA-MMoE}(\text{RMSNorm}(x_{i,m}^{l-1})), \\ x_{i,m}^l &= x_{i,m}^{l'} + \text{FFN-MMoE}(\text{RMSNorm}(x_{i,m}^{l'})), \end{aligned} \tag{1}$$

¹Due to limited computational resource and time, current version of NaViL-9B in this paper is only trained with 450B tokens.

Table 2: **Comparison between NaViL-9B and existing MLLMs on multimodal benchmarks.** “#A-Param” denotes the number of activated parameters. [†]InternVL-2.5-8B adopts the same high-quality data with NaViL-9B, so we mark it as the compositional counterpart. Note that its 300M visual encoder is distilled from another 6B large encoder. **Bold** and underline indicate the best and the second-best performance among native MLLMs, respectively. For MME, we sum the perception and cognition scores. Average scores are computed by normalizing each metric to a range between 0 and 100.

Model	#A-Param	Avg	MMVet	MMMU	MMB	MME	MathVista	OCR-B	TVQA	DocVQA	AI2D	ChartQA	InfoVQA
<i>Compositional MLLMs:</i>													
MobileVLM-V2 [10]	1.7B	—	—	—	57.7	—	—	—	—	—	—	—	—
MobileVLM-V2 [10]	3.0B	—	—	—	63.2	—	—	—	57.5	—	—	—	—
Mini-Gemini [15]	3.5B	—	31.1	31.7	59.8	1653	29.4	—	56.2	34.2	—	—	—
MM1-MoE-Chat [18]	3.5B	—	42.2	38.6	70.8	1772	32.6	—	72.9	—	—	—	—
DeepSeek-VL [16]	2.0B	—	34.8	32.2	64.6	1532	31.1	409	57.8	—	51.5	—	—
PaliGemma [3]	2.9B	—	33.1	34.9	71.0	1686	28.7	614	68.1	—	68.3	—	—
MiniCPM-V-2 [24]	2.8B	—	41.0	38.2	69.1	1809	38.7	605	74.1	71.9	62.9	—	—
InternVL-1.5 [9]	2.2B	61.3	39.3	34.6	70.9	1902	41.1	654	70.5	85.0	69.8	74.8	55.4
Qwen2VL [22]	2.1B	67.3	49.5	41.1	74.9	1872	43.0	809	79.7	90.1	74.7	73.5	65.5
InternVL-2.5 [8]	2.2B	69.6	60.8	43.6	74.7	2138	51.3	804	74.3	88.7	74.9	79.2	60.9
Qwen2VL [22]	8.2B	77.1	62.0	54.1	83.0	2327	58.2	866	84.3	94.5	83.0	83.0	76.5
Qwen2.5-VL [1]	8.2B	80.2	67.1	58.6	83.5	2347	68.2	864	84.9	95.7	83.9	87.3	82.6
[†] InternVL-2.5 [8]	8.1B	77.3	62.8	56.0	84.6	2344	64.4	822	79.1	91.9	84.5	84.8	75.7
<i>Native MLLMs:</i>													
Fuyu-8B (HD) [2]	8B	—	21.4	—	10.7	—	—	—	—	—	64.5	—	—
SOLO [6]	7B	—	—	—	—	1260	34.4	—	—	—	61.4	—	—
Chameleon-7B ² [5]	7B	14.0	8.3	25.4	31.1	170	22.3	7	4.8	1.5	46.0	2.9	5.0
EVE-7B [12]	7B	34.6	25.6	32.3	49.5	1483	25.2	327	51.9	22.0	48.5	19.5	20.0
EVE-7B (HD) [12]	7B	45.2	25.7	32.6	52.3	1628	34.2	398	56.8	53.0	61.0	59.1	25.0
Emu3 [23]	8B	—	37.2	31.6	58.5	—	—	687	64.7	76.3	70.0	68.6	43.8
VoRA [21]	7B	—	33.7	32.2	64.2	1674	—	—	56.3	—	65.6	—	—
VoRA-AnyRes [21]	7B	—	33.7	32.0	61.3	1655	—	—	58.7	—	61.1	—	—
EVEv2 [13]	7B	—	45.0	39.3	66.3	1709	—	702	71.1	—	74.8	73.9	—
SAIL [14]	7B	—	46.3	—	70.1	1719	<u>57.0</u>	783	<u>77.1</u>	—	<u>76.7</u>	—	—
Mono-InternVL [17]	1.8B	60.6	40.1	33.7	65.5	<u>1875</u>	45.7	767	72.6	80.0	68.6	73.7	43.0
NaViL-2B (ours)	2.4B	<u>68.8</u>	<u>78.3</u>	<u>41.8</u>	<u>71.2</u>	1822	50.0	796	76.9	<u>85.4</u>	<u>74.6</u>	<u>78.0</u>	<u>56.0</u>
NaViL-9B (ours)	9.2B	77.0	79.6	54.7	76.5	2225	66.7	837	77.2	90.6	82.4	85.4	70.2

where $\text{RMSNorm}(\cdot)$ is the layer normalization operation, and $\text{MHA-MMoE}(\cdot)$ and $\text{FFN-MMoE}(\cdot)$ are the modality-specific attention and FFN expert, respectively, formulated by

$$\begin{aligned}
\text{MHA-MMoE}(x_{i,m}) &= (\text{softmax}(\frac{QK^T}{\sqrt{d}})V)W_O^m, \\
Q_{i,m} &= x_{i,m}W_Q^m, K_{i,m} = x_{i,m}W_K^m, V_{i,m} = x_{i,m}W_V^m, \\
\text{FFN-MMoE}(x_{i,m}) &= (\text{SiLU}(x_{i,m}W_{\text{gate}}^m) \odot x_{i,m}W_{\text{up}}^m)W_{\text{down}}^m.
\end{aligned} \tag{2}$$

Here $W_Q^m, W_K^m, W_V^m, W_O^m$ and $W_{\text{gate}}^m, W_{\text{up}}^m, W_{\text{down}}^m$ are all modality-specific projection matrices, and $\text{SiLU}(\cdot)$ denotes the activation function, \odot denotes the element-wise product operation.

D The NLP capability

We also evaluate the NLP capability of our model on three popular NLP tasks, as shown in Tab. 5. Thanks to the modality-specific MoE architecture, NaViL maintains the NLP capabilities of its initialization LLM (Qwen3-8B). Despite not using a large amount of high-quality text data, NaViL performs well on the common NLP tasks and show much stronger NLP capabilities compared to other native MLLMs, showing its data efficiency.

²The performance of Chameleon-7B is from [17].

Table 3: **Hyper-parameters for training NaViL-2B.**

Configuration	Multi-modal Generative Pre-training (S1)		Supervised Fine-tuning (S2)
	S1.1	S1.2	
Maximum number of image patches	4096	12188	24576
Training steps	70k	40k	30k
Global batch size	7,000	4,614	2,234
Weight decay	0.05	0.1	0.01
Learning rate schedule	constant with warm-up		cosine decay
Peak learning rate	$5e^{-5}$		$2e^{-5}$
Visual Multi-scale Packing	✓		
LLM max sequence length	16,384		
Warm-up steps	200		
Optimizer	AdamW		
Optimizer hyperparameters	$\beta_1 = 0.9, \beta_2 = 0.95, eps = 1e^{-8}$		
Gradient accumulation	1		
Numerical precision	bfloat16		

Table 4: **Hyper-parameters for training NaViL-9B.**

Configuration	Multi-modal Generative Pre-training (S1)		Supervised Fine-tuning (S2)
	S1.1	S1.2	
Maximum number of image patches	4096	12188	24576
Training steps	50k	33k	6k
Weight decay	0.05	0.1	0.01
Global batch size	10,300	1,792	3,520
Visual Multi-scale Packing	✗	✓	✓
Learning rate schedule	constant with warm-up		cosine decay
Peak learning rate	$5e^{-5}$		$2e^{-5}$
LLM max sequence length	16,384		
Warm-up steps	200		
Optimizer	AdamW		
Optimizer hyperparameters	$\beta_1 = 0.9, \beta_2 = 0.95, eps = 1e^{-8}$		
Gradient accumulation	1		
Numerical precision	bfloat16		

Table 5: **Comparison of NaViL and existing native MLLMs on three common NLP tasks.** Except for Chameleon, models are evaluated using OpenCompass toolkit [11].

Models	#A-Param	MMLU	CMMLU	MATH
InternLM2-Chat [19]	1.8B	47.1	46.1	13.9
Qwen3-8B (non-thinking) [20]	8B	76.5	76.8	71.1
EVE [12]	7B	43.9	33.4	0.7
Chameleon [5]	7B	52.1	-	11.5
Mono-InternVL [17]	2B	45.1	44.0	12.3
NaViL-9B (ours)	9.2B	74.9	75.1	66.2

41 E More Qualitative Results

42 More visualization results of multimodal understanding are provided below.

References

- [1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.
- [2] Rohan Bavishi, Erich Elsen, Curtis Hawthorne, Maxwell Nye, Augustus Odena, Arushi Somani, and Sağnak Taşlılar. Introducing our multimodal models, 2023.
- [3] Lucas Beyer, Andreas Steiner, André Susano Pinto, Alexander Kolesnikov, Xiao Wang, Daniel Salz, Maxim Neumann, Ibrahim Alabdulmohsin, Michael Tschannen, Emanuele Bugliarello, et al. Paligemma: A versatile 3b vlm for transfer. *arXiv preprint arXiv:2407.07726*, 2024.
- [4] Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, et al. Internlm2 technical report. *arXiv preprint arXiv:2403.17297*, 2024.
- [5] ChameleonTeam. Chameleon: Mixed-modal early-fusion foundation models. *arXiv preprint arXiv:2405.09818*, 2024.
- [6] Yangyi Chen, Xingyao Wang, Hao Peng, and Heng Ji. A single transformer for scalable vision-language modeling. *arXiv preprint arXiv:2407.06438*, 2024.
- [7] Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, 2024.
- [8] Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, 2024.
- [9] Zhe Chen, Weiyun Wang, Hao Tian, Shenglong Ye, Zhangwei Gao, Erfei Cui, Wenwen Tong, Kongzhi Hu, Jiapeng Luo, Zheng Ma, et al. How far are we to gpt-4v? closing the gap to commercial multimodal models with open-source suites. *arXiv:2404.16821*, 2024.
- [10] Xiangxiang Chu, Limeng Qiao, Xinyu Zhang, Shuang Xu, Fei Wei, Yang Yang, Xiaofei Sun, Yiming Hu, Xinyang Lin, Bo Zhang, et al. Mobilevlm v2: Faster and stronger baseline for vision language model. *arXiv preprint arXiv:2402.03766*, 2024.
- [11] Contributors. Opencompass: A universal evaluation platform for foundation models. <https://github.com/open-compass/opencompass>, 2023.
- [12] Haiwen Diao, Yufeng Cui, Xiaotong Li, Yueze Wang, Huchuan Lu, and Xinlong Wang. Unveiling encoder-free vision-language models. *arXiv preprint arXiv:2406.11832*, 2024.
- [13] Haiwen Diao, Xiaotong Li, Yufeng Cui, Yueze Wang, Haoge Deng, Ting Pan, Wenxuan Wang, Huchuan Lu, and Xinlong Wang. Evev2: Improved baselines for encoder-free vision-language models. *arXiv preprint arXiv:2502.06788*, 2025.
- [14] Weixian Lei, Jiacong Wang, Haochen Wang, Xiangtai Li, Jun Hao Liew, Jiashi Feng, and Zilong Huang. The scalability of simplicity: Empirical analysis of vision-language learning with a single transformer. *arXiv preprint arXiv:2504.10462*, 2025.
- [15] Yanwei Li, Yuechen Zhang, Chengyao Wang, Zhisheng Zhong, Yixin Chen, Ruihang Chu, Shaoteng Liu, and Jiaya Jia. Mini-gemini: Mining the potential of multi-modality vision language models. *arXiv:2403.18814*, 2024.
- [16] Haoyu Lu, Wen Liu, Bo Zhang, Bingxuan Wang, Kai Dong, Bo Liu, Jingxiang Sun, Tongzheng Ren, Zhuoshu Li, Yaofeng Sun, et al. Deepseek-vl: Towards real-world vision-language understanding. *arXiv preprint arXiv:2403.05525*, 2024.
- [17] Gen Luo, Xue Yang, Wenhan Dou, Zhaokai Wang, Jiawen Liu, Jifeng Dai, Yu Qiao, and Xizhou Zhu. Mono-intervl: Pushing the boundaries of monolithic multimodal large language models with endogenous visual pre-training. In *CVPR*, 2025.
- [18] Brandon McKinzie, Zhe Gan, Jean-Philippe Fauconnier, Sam Dodge, Bowen Zhang, Philipp Dufter, Dhruvi Shah, Xianzhi Du, Futang Peng, Floris Weers, Anton Belyi, Haotian Zhang, Karanjeet Singh, Doug Kang, Ankur Jain, Hongyu He, Max Schwarzer, Tom Gunter, Xiang Kong, Aonan Zhang, Jianyu Wang, Chong Wang, Nan Du, Tao Lei, Sam Wiseman, Guoli Yin, Mark Lee, Zirui Wang, Ruoming Pang, Peter Grasch, Alexander Toshev, and Yinfei Yang. MM1: methods, analysis & insights from multimodal LLM pre-training. *arXiv:2403.09611*, 2024.

- 94 [19] InternLM Team. Internlm: A multilingual language model with progressively enhanced capabilities.
95 <https://github.com/InternLM/InternLM>, 2023.
- 96 [20] Qwen Team. Qwen3 blog. <https://qwenlm.github.io/blog/qwen3/>, 2025.
- 97 [21] Han Wang, Yongjie Ye, Bingru Li, Yuxiang Nie, Jinghui Lu, Jingqun Tang, Yanjie Wang, and Can Huang.
98 Vision as lora. *arXiv preprint arXiv:2503.20680*, 2025.
- 99 [22] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin
100 Wang, Wenbin Ge, et al. Qwen2-vl: Enhancing vision-language model’s perception of the world at any
101 resolution. *arXiv preprint arXiv:2409.12191*, 2024.
- 102 [23] Xinlong Wang, Xiaosong Zhang, Zhengxiong Luo, Quan Sun, Yufeng Cui, Jinsheng Wang, Fan Zhang,
103 Yueze Wang, Zhen Li, Qiyang Yu, Yingli Zhao, Yulong Ao, Xuebin Min, Tao Li, Boya Wu, Bo Zhao,
104 Bowen Zhang, Liangdong Wang, Guang Liu, Zheqi He, Xi Yang, Jingjing Liu, Yonghua Lin, Tiejun Huang,
105 and Zhongyuan Wang. Emu3: Next-token prediction is all you need. *arXiv: 2409.18869*, 2024.
- 106 [24] Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin
107 Zhao, Zhihui He, et al. Minicpm-v: A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*,
108 2024.

Image Captioning and Visual Question Answering



Prompt: Describe this image in detail.

Response: The image shows an aerial view of a modern architectural complex situated near a coastal area. The building has a distinctive design with a series of curved, wave-like structures on its roof, giving it a futuristic appearance. The exterior is primarily composed of light-colored materials, possibly concrete or stone, with large glass windows.

In front of the building, there is a wide, multi-lane road lined with palm trees, suggesting a warm, coastal climate. The road appears to be a major thoroughfare, with designated bike lanes and pedestrian pathways. There are also train tracks running parallel to the road, indicating the presence of a public transportation system.

The background features a large body of water, likely a bay or ocean, with several boats and yachts docked at a marina. The coastline is lined with more buildings, and there are hills or mountains visible in the distance, adding to the scenic view. The overall setting suggests a bustling urban area with a focus on modern infrastructure and a strong connection to the waterfront.

Prompt: Which city is this building most likely in?

Response: This building is most likely in San Diego, California. The architectural style and the surrounding landscape, including the water and palm trees, are characteristic of San Diego.



Prompt: Who is the person in black in the middle?

Response: The person in black in the middle is Lionel Messi.

Prompt: What are they celebrating about?

Response: The image shows a group of soccer players celebrating with a trophy, wearing medals and jerseys. They are likely celebrating a victory in a soccer tournament or championship.



Prompt: What is the content in this image?

Response: The image appears to be a screenshot from a video game, likely Grand Theft Auto V (GTA V), given the visual style and interface elements. It shows a character standing on a sidewalk in a small town or suburban area. The environment includes a street with tram tracks, a few buildings with signs, and some cars parked along the road. The buildings have signs for businesses like "Wine" and "On The Bone." The scene has a somewhat deserted look, with no other characters or significant activity visible.

OCR

NUTRITIONAL INFORMATION		
Serving size: 10 g		Servings per package: 25
	Average Quantity Per Serving	Average Quantity Per 100g
Energy	143kJ	1430 kJ
Protein	less than 0.1g	0.16g
Fat, total	less than 0.1g	0.37g
- saturated	less than 0.1g	0.26g
Carbohydrates	8.31g	83.1g
Sugars	8.11g	81.1g
Sodium	0.41mg	4.04mg
INGREDIENTS: 100% Pure New Zealand Honey		
Keep Stored Below 20 deg°C, BPA-Free Container.		
Not suitable for children under 1 year of age.		

Prompt: Extract all the text in this image.

Response: NUTRITIONAL INFORMATION

Serving size: 10 g

Servings per package: 25

Average Quantity Per Serving | Average Quantity Per 100g

Energy | 143kJ | 1430 kJ

Protein | less than 0.1g | 0.16g

Fat, total | less than 0.1g | 0.37g

- saturated | less than 0.1g | 0.26g

Carbohydrates | 8.31g | 83.1g

Sugars | 8.11g | 81.1g

Sodium | 0.41mg | 4.04mg

INGREDIENTS: 100% Pure New Zealand Honey

Keep Stored Below 20 deg°C, BPA-Free Container.

Not suitable for children under 1 year of age.

Food = Lunch x 5 people

CHAPELBANK HOTEL
69 EAST HIGH STREET
FORFAR DD8 2EP
TEL NO 01307 463151

TABLE 5
RESTAURANT

3x SOUP	3.95	11.85
Lunch Main		
Lasagne (V)		6.95
MISC FOOD		6.95
2x CHIPS / ONION RINGS	1.75	3.50
-FOOD		29.25--
Prev Bal:		29.25
SUBTOTAL		29.25

Prompt: Describe the image.

Response: The image is a receipt from Chapelbank Hotel for a lunch order for five people. The items listed on the receipt include 3x soup at £3.95 each, totaling £11.85; lunch main course at £6.95; lasagne (V) at £6.95; misc food at £6.95; and 2x chips/onion rings at £1.75 each, totaling £3.50. The subtotal for the food is £29.25. The receipt also includes a previous balance of £29.25, making the total amount due £29.25. The hotel's address is 69 East High Street, Forfar DD8 2EP, and the telephone number is 01307 463151.

Prompt: How many people are attending this lunch?

Response: There are 5 people attending this lunch. This is indicated at the top of the receipt where it says "Food = Lunch x 5 people".



Prompt: Where is the left exit going to?

Response: The left exit is going to Reston, Herndon, and Dulles Airport.

Prompt: Which road leads to the EXIT 46 B?

Response: The road that leads to EXIT 46 B is Chain Br Rd North.

Multi-lingual OCR

作文(60分)

题目 专注目标, 享受生命.

目标之光, 如月光清辉; 生命之辉, 如群星闪耀。失去了群星, 月色便清冷孤寂; 不见了明月, 群星也势必黯然失色。只有专注于月光, 方有方向; 只有欣赏群星, 方可享受生命。星月交辉, 夜空撩人; 专注目标, 享受生命, 方可体味人间清欢。

Prompt: 请先提取里面的文本, 再翻译为英文。

Response:

文本内容:

作文 (60分) 题目: 专注目标, 享受生命。

目标之光, 如月光清辉; 生命之辉, 如群星闪耀。失去了群星, 月色便显得冷寂; 不见了明月, 群星也势必黯然失色。只有专注于月光, 方有方向; 只有欣赏群星, 方可享受生命。星月交辉, 夜空撩人; 专注目标, 享受生命, 方可体味人间清欢。

翻译为英文:

Essay (60 points)

Title: Focus on your goal, enjoy life.

The light of the goal is like the clear glow of moonlight; the brilliance of life is like the dazzling stars. Without the stars, the moonlight becomes cold and desolate; without the moon, the stars will inevitably become dull. Only by focusing on the moonlight can you find direction; only by appreciating the stars can you enjoy life. The stars and moon shine together, making the night sky enchanting; focusing on your goal, enjoying life, allows you to savor the quiet joy of life.

Document Understanding

KBA MARKETING GROUP — 2001 VENUE SPONSORSHIP AGREEMENT

THIS VENUE SPONSORSHIP AGREEMENT (the "Agreement") is entered into as of the following Date of Agreement, by and between KBA Marketing Group ("KBA") and the owner ("Venue Owner") of the business establishment ("Venue") described below. Below are various defined terms for the purposes of this Agreement.

Date of Agreement	3/3/01
Term of Agreement	January 1, 2001 — December 31, 2001
Sponsorship Program	Camel Club Program ("CCP")

I. VENUE INFORMATION

Venue Name	Vitucci's	Venue Code	438
Venue Address	1832 East North Avenue		
City/State/Zip Code	Milwaukee WI 53202		
Phone no. (XXX-XXX-XXXX)	PERIODICALLY CONFIDENTIAL MATERIAL REDACTED		
Venue Owner Name	Vitucci's, Inc.		
Venue Owner Category	Corporation		

II. SPONSORSHIP FUNDS AND EXTRA VALUE ITEMS

Special Events Fund: \$1,000.00 to offset costs for a minimum of 3 Special Events from Jan. 1 — June 30, 2001.
\$1,000.00 to offset costs for a minimum of 3 Special Events from July 1 — Dec. 31, 2001.

Printed Materials Fund: \$100.00 to offset costs for the printing and display of certain items, as well as the purchase of local media, including quarter-page advertisements, and additional printed material and display advertising, as available, at the discretion of KBA, from Jan. 1 — Dec. 31, 2001.

Total Direct Payments: \$2,000.00, payable on or before the following payment dates in the following amounts:
March 31, 2001: \$500.00; June 30, 2001: \$500.00;
September 30, 2001: \$500.00; December 31, 2001: \$500.00.

TOTAL SPONSORSHIP FUNDS: \$4,000.00 [This amount includes all sub-totals set forth above]

KBA will also provide the following items, value items at no charge to Venue:

Excluded License Fees: All state and local retail tobacco licensing fees for the Venue (if Venue is currently without a license), plus assistance with tobacco licensing application procedures.

Bar Essentials: Branded ashtrays, beverage napkins, generic branded matchbooks, and other bar essentials delivered to Venue for use within the Venue, as available, at the discretion of KBA.

Custom Matchbooks: 3 cases of custom branded matchbooks that display the Venue's name and logo, delivered to Venue.

Flyer Distribution: Public distribution of flyers for Special Events and activities at Venue, at discretion of KBA.

Cigarette Delivery: Free delivery of Camel cigarette brand styles (the "Product") purchased by Venue Owner for resale to Venue. Free (On-site) Product that KBA will distribute to Venue to adult smokers (as permitted by law).

COPY ORIGINAL

Initiate: KBA *[Signature]* Venue Owner *[Signature]*

Source: <https://www.industrydocuments.ucsf.edu/docs/fmj0226>

Prompt: What is the duration of this venue sponsorship agreement?

Response: January 1, 2001 - December 31, 2001

Prompt: Where is Vitucci's located (complete address)?

Response: 1832 East North Avenue, Milwaukee WI 53202

Prompt: What cigarette brand is mentioned in the agreement for delivery and distribution?

Response: Camel cigarette brand styles purchased by Venue Owner for resale in Venue

Prompt: Who has discretion over the distribution of flyers for special events?

Response: KBA Marketing Group ("KBA") has discretion over the distribution of flyers for special events.

LaTeX and Math

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}[q \sim P(Q), \{o_i\}_{i=1}^G \sim \pi_{\theta_{old}}(O|q)]$$

$$\frac{1}{G} \sum_{i=1}^G \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \left\{ \min \left[\frac{\pi_{\theta}(o_{i,t}|q, o_{i,<t})}{\pi_{\theta_{old}}(o_{i,t}|q, o_{i,<t})} \hat{A}_{i,t}, \text{clip} \left(\frac{\pi_{\theta}(o_{i,t}|q, o_{i,<t})}{\pi_{\theta_{old}}(o_{i,t}|q, o_{i,<t})}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}_{i,t} \right] - \beta \mathbb{D}_{KL} [\pi_{\theta} || \pi_{ref}] \right\}, \quad (3)$$

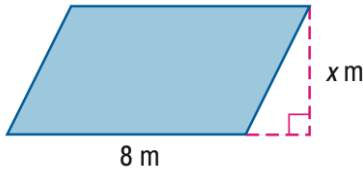
Prompt: Convert the formula in the figure to latex format.

Response:

```

“““\latex
\begin{equation*}
\begin{split}
&\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}[\mathbb{q} \sim P(Q), \{o_i\}_{i=1}^G \\
&\sim \pi_{\theta_{old}}(O|\mathbb{q})] \\
&\frac{1}{G} \sum_{i=1}^G \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \left\{ \min \left[ \frac{\pi_{\theta}(o_{i,t}|\mathbb{q}, o_{i,<t})}{\pi_{\theta_{old}}(o_{i,t}|\mathbb{q}, o_{i,<t})} \hat{A}_{i,t}, \text{clip} \left( \frac{\pi_{\theta}(o_{i,t}|\mathbb{q}, o_{i,<t})}{\pi_{\theta_{old}}(o_{i,t}|\mathbb{q}, o_{i,<t})}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}_{i,t} \right] - \beta \mathbb{D}_{KL} [\pi_{\theta} || \pi_{ref}] \right\}, \\
&\end{split}
\end{equation*}
”””

```



Prompt: Find x . $A = 148 \text{ m}^2$. Please think step by step.

Response: To find the value of x in the parallelogram, we can use the formula for the area of a parallelogram:

$$A = \text{base} \times \text{height}$$

Given: - Area $A = 148 \text{ m}^2$ - Base = 8 m - Height = x m
Substitute the given values into the formula:

$$148 = 8 \times x$$

Solve for x :

$$x = \frac{148}{8}$$

$$x = 18.5 \text{ m}$$

So, $x = 18.5 \text{ m}$.

Weight of winning dog at a dog show	
Year	Weight (kg)
2013	27
2014	28
2015	34
2016	29
2017	40

Prompt: A dog show enthusiast recorded the weight of the winning dog at recent dog shows. According to the table, what was the rate of change between 2016 and 2017? (Unit: kilograms per year)

Response: To determine the rate of change between 2016 and 2017, we use the formula for rate of change:

$$\text{Rate of Change} = \frac{\text{Weight in 2017} - \text{Weight in 2016}}{\text{Year 2017} - \text{Year 2016}}$$

From the table: - Weight in 2016 = 29 kg - Weight in 2017 = 40 kg

Substitute these values into the formula:

$$\text{Rate of Change} = \frac{40 \text{ kg} - 29 \text{ kg}}{2017 - 2016} = \frac{11 \text{ kg}}{1 \text{ year}}$$

$$\text{Rate of Change} = 11 \text{ kg/year}$$

Thus, the rate of change between 2016 and 2017 is 11 kg/year.