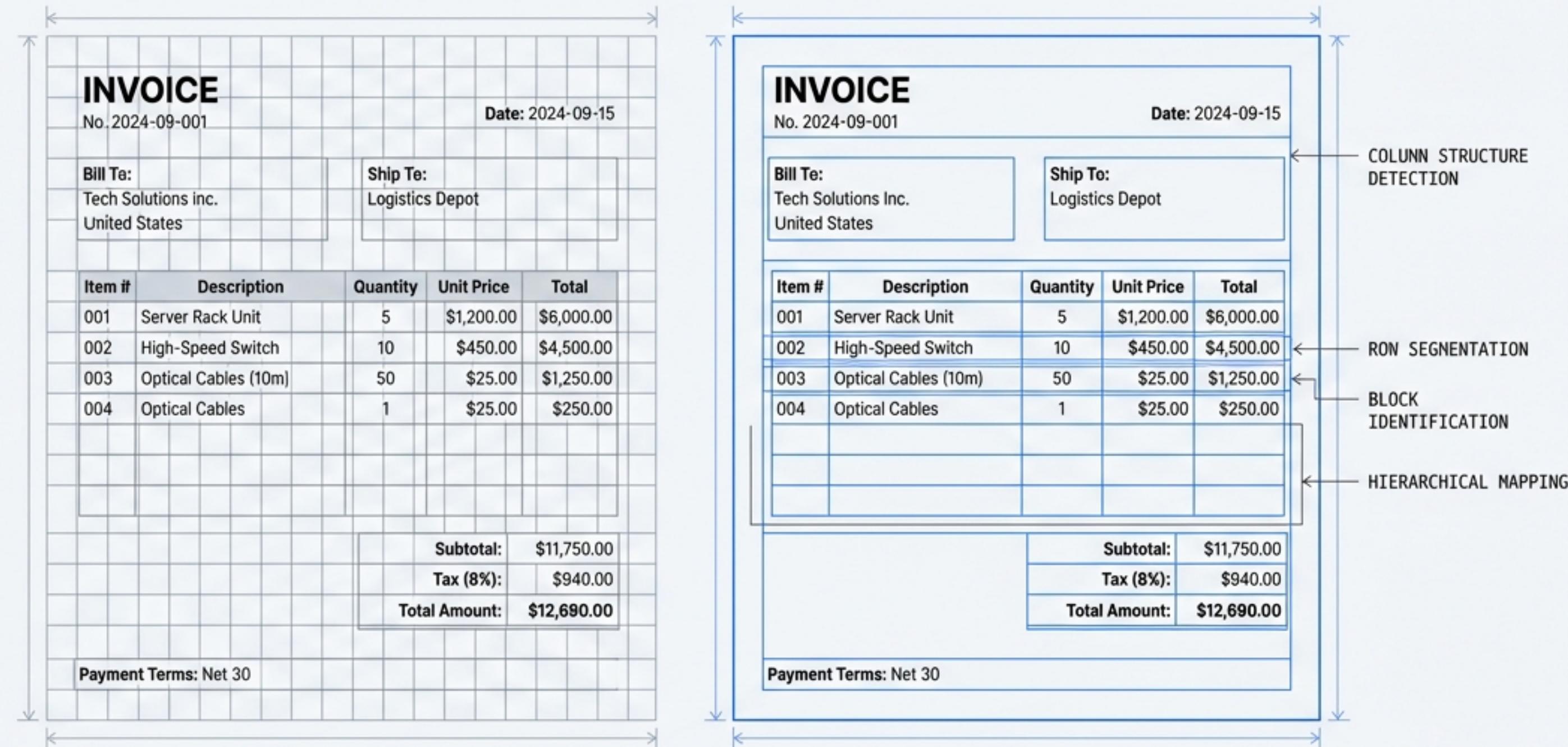


# 문서 중심 멀티모달 모델을 위한 구조 분해 어텐션

자연 이미지 인코더의 한계를 넘어선 구조적 귀납 편향(Structural Inductive Bias)의 주입

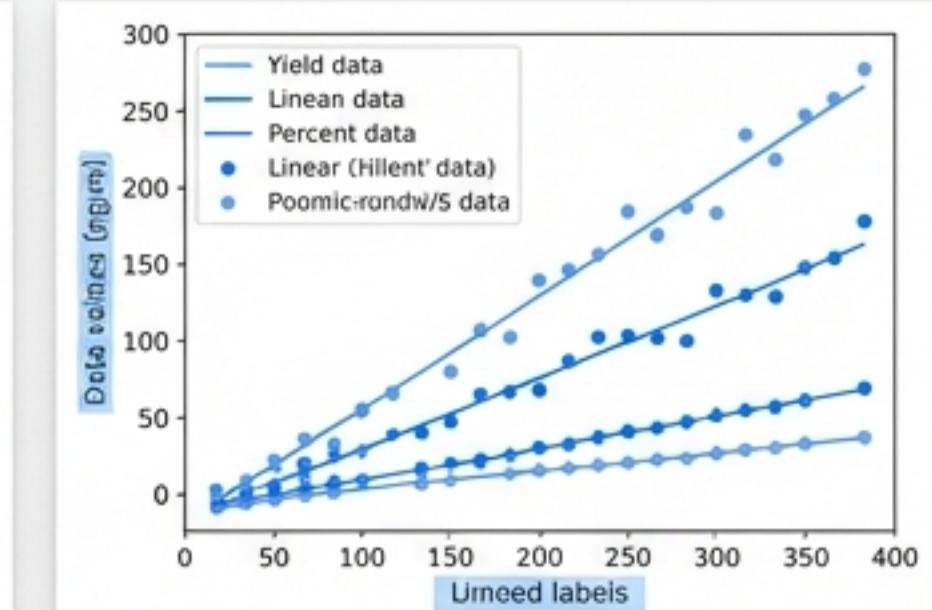
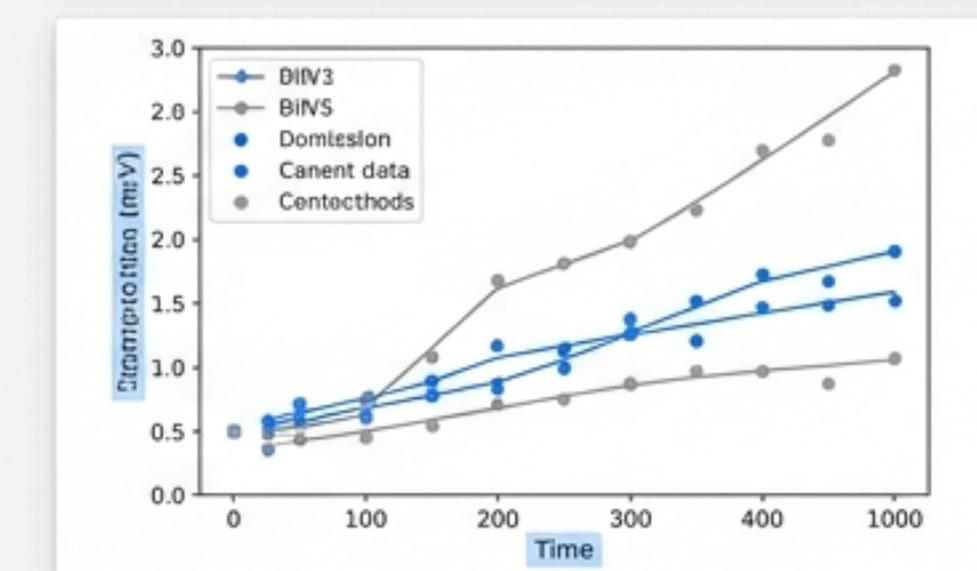


# 문서 이해(Document Understanding)에 대한 수요 급증

ChartQA, DocVQA와 같은 복잡한 문서 처리 작업에서 MLLM의 중요성이 커지고 있습니다.

기존 모델들(InternVL, LLaVA 등)은 강력하지만, 여전히 '자연 이미지' 처리를 위한 설계된 비전 인코더에 의존하고 있습니다.

	Balance Sheet		Comparisons & Statement	
	2021	2023	2021	2023
<b>ASSETS</b>				
Non-current assets				
Goodwill and intangible assets	\$ 5,566,000	\$ 4,806,870	\$ 5,050,000	\$ 5,328,000
Property, plant and equipment	1,866,900	226,200	1,200,000	1,100,200
Investments in associates	905,900	903,300	—	—
Other assets	546,900	369,200	365,900	346,900
Other non-current assets	30,000	30,000	200,200	24,000
Total assets	<b>\$ 10,606,870</b>	<b>\$ 5,506,270</b>	<b>\$ 10,566,200</b>	<b>\$ 5,328,000</b>
Current assets				
Cash and cash equivalents	550,370	399,200	266,800	522,300
Other current assets				
Total assets	<b>\$ 10,606,870</b>	<b>\$ 5,506,270</b>	<b>\$ 10,566,200</b>	<b>\$ 5,328,000</b>
<b>TOTAL ASSETS</b>	<b>\$ 10,606,870</b>	<b>\$ 5,506,270</b>	<b>\$ 10,566,200</b>	<b>\$ 5,328,000</b>
<b>Debtors and receivables</b>				
Total trade debtors	1,141,640	1,200,960	641,511	536,200
Trade receivable	—	33,900	—	33,219
Other debtors from customers	(25,135)	300	(39,000)	23,305
Other receivables and advances	120,450	69,600	—	11,600
Other creditors	150,400	611,700	260,000	196,300
Total debtors	<b>\$ 1,116,050</b>	<b>\$ 1,200,960</b>	<b>\$ 1,210,060</b>	<b>\$ 1,166,810</b>
Indirectors' contributions	316,790	256,300	72,291	101,310
Other amounts receivable	—	—	(52,205)	—
Other debts and receivables			20,000	1,007
<b>TOTAL DEBTORS</b>	<b>\$ 1,442,790</b>	<b>\$ 1,200,960</b>	<b>\$ 1,288,356</b>	<b>\$ 1,166,810</b>
<b>Financial interests</b>				
Liquid investments	2,346,853	2,430,370	2,275,063	2,322,453
Other financial assets	—	—	260,995	101,000
<b>TOTAL FINANCIAL INTERESTS</b>	<b>\$ 2,346,853</b>	<b>\$ 2,430,370</b>	<b>\$ 2,536,058</b>	<b>\$ 2,423,453</b>
<b>Debtors</b>				
Financial assets:				
Financial assets	1,038,475	1,005,950	1,022,575	1,105,310
Unused one-off pre-completion advances	—	(50,000)	—	802,905
Creditors from clients	430,400	—	(10,000)	—
Other receivables and advances	28,225	62,600	—	—
Other debts	—	601,700	225,000	414,307
Total debtors	<b>\$ 1,606,600</b>	<b>\$ 1,100,270</b>	<b>\$ 1,260,200</b>	<b>\$ 1,005,310</b>
Other receivables	60,000	37,900	—	100,000
<b>TOTAL DEBTORS</b>	<b>\$ 1,666,600</b>	<b>\$ 1,138,170</b>	<b>\$ 1,260,200</b>	<b>\$ 1,005,310</b>
<b>Expenditure capitalised</b>				
Total assets	<b>\$ 10,606,870</b>	<b>\$ 5,506,270</b>	<b>\$ 10,566,200</b>	<b>\$ 5,328,000</b>
Capitalised provisions	480,950	247,275	506,067	466,298
Provisions	20,581	10,680	22,961	10,000
Protected assets	2,050	(1,879)	(1,900)	—
Total reserves	600,500	1,200,645	1,360,200	1,906,567
<b>Total assets</b>	<b>\$ 10,606,870</b>	<b>\$ 5,506,270</b>	<b>\$ 10,566,200</b>	<b>\$ 5,328,000</b>



A Recount of Commutias Academic Papers

Sennaroglu, Wittenberg, Shafiq, Normoyle and Burnstock: *Incora Potentia*

**Illustration 1:** *Conseil des experts* (*Commission d'experts*) dans les domaines de l'écologie et de la santé publique. Les deux commissions sont chargées de faire des recommandations sur les meilleures stratégies pour assurer la sécurité et la santé publique. Elles peuvent recommander des politiques et des mesures réglementaires ou législatives, ou faire des recommandations sur les meilleures stratégies pour assurer la sécurité et la santé publique. Elles peuvent recommander des politiques et des mesures réglementaires ou législatives, ou faire des recommandations sur les meilleures stratégies pour assurer la sécurité et la santé publique.

**Благодаря** (запомнил падение времени) **они** (запомнили)

**I** «muri» di difesa possono essere sia in conformità che a distanza, oppure difensivi unicamente. Questi termini sono generalmente riferiti ai lavori del National Guard, ma i diversi eserciti spesso distinguono le imprese dei loro contingenti da quelli dei contingenti controllati anche congiuntamente con il resto dell'esercito.

Les deux dernières années ont été marquées par une croissance de l'industrie et de la construction dans le secteur manufacturier et dans les services.

Все эти факторы, конечно, должны учитываться в процессе разработки стратегии продвижения. Но в конечном итоге, для успешного продвижения на рынке необходимо определить, каким образом можно привлечь внимание потребителей к новому продукту.

Tanque s'assentant sur l'application de la méthode des dérivées partielles pour la détermination des courants dans les réseaux électriques, cette théorie est basée sur l'hypothèse que les courants peuvent être exprimés sous la forme d'une somme de courants permanents et temporaires.

Важен етът етап във възстановяването на здравето и възможността за възстановяване на работната способност.

	DCT	DMT	DMT	DMT
Welding	0.748	0.748	0.748	0.748
Welding	14.83	14.83	14.83	14.83
Electrode Welding	2.03	1.018	0.6000	0.3000

$$R_{\text{eff}} = \frac{\text{flat} - \text{fit}}{\text{flat} - \text{min}}$$

DREIF - 1960

Ваша команда: **Городской патруль**

Следует отметить, что в ходе заседания комиссии по вопросам правоприменения в сфере труда и социальной политики в Краснодаре было подтверждено, что в Краснодарском крае не выявлено фактов нарушения прав работников в сфере труда.

жълт кабинет имат широка външна архитектура, но същевременно  
имат различни форми на изходи и коридори.

Ensuite, l'heure est venue de faire évoluer les deux dernières parties de la loi pour faire émerger une nouvelle réglementation qui sera plus adaptée à la situation actuelle.

đã là một khía cạnh quan trọng trong việc xác định giá trị của một doanh nghiệp.

и в то же время для этого и выразил все вышеупомянутые виды субстанций.

Wie darf man selbst aktives Elemente in älteren Werkkulturen als Resonanz

Published 2003 generation Customer Information

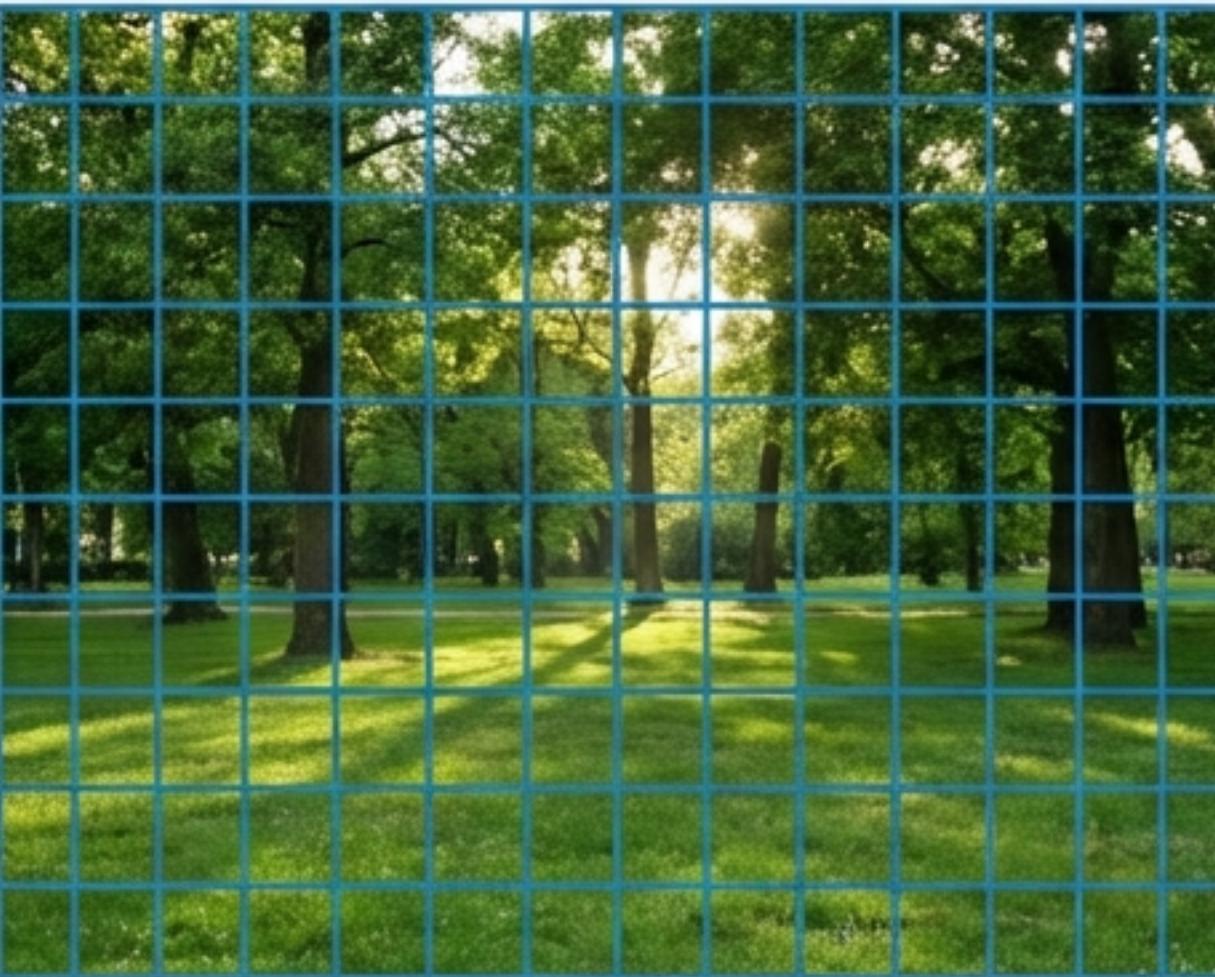
—**W**ILHELMUS VAN DER HORST, 1900-1901, was born at Utrecht, Holland, on June 10, 1868.

the heart of each pair consists of longitudinal fibers that run parallel to the surface, while the outermost layer consists of transverse fibers.

www.elsevier.com/locate/jtbi

# 근본적인 설계 불일치: 자연 이미지 vs. 구조화된 문서

자연 이미지 (Natural Image)



연속적 정보 (Continuous Info)

픽셀 간 높은 상관관계, 균일한 패치 처리 가능.

문서 (Document)

INVOICE					Date: 10. 2023																								
Ferry Finance Naaren patress. 129HOB alt soem octation 125000 Ord. No001			Billed by Author Santions Searily: Stase Orde Elevmout Citlom																										
Classification: <input type="checkbox"/> Male <input type="checkbox"/> Male <input type="checkbox"/> Pcoale <input type="checkbox"/> Other																													
Rechnungserstellung (pure-neatiod) <input type="checkbox"/> Minus Stotet <input type="checkbox"/> Suppem amption/entment																													
<table><thead><tr><th>#</th><th>Description</th><th>Qty</th><th>Net</th><th>Total</th><th>Amt/Unit</th></tr></thead><tbody><tr><td>1</td><td>Slock to see Romix</td><td>1</td><td>1</td><td>260.00</td><td>260.00</td></tr><tr><td>2</td><td>Micropla tabocat 1000</td><td>1</td><td>1</td><td>150.00</td><td>150.00</td></tr><tr><td>3</td><td>Ocensm oten lor a emennd</td><td>2</td><td>1</td><td>36.00</td><td>36.00</td></tr></tbody></table>			#	Description	Qty	Net	Total	Amt/Unit	1	Slock to see Romix	1	1	260.00	260.00	2	Micropla tabocat 1000	1	1	150.00	150.00	3	Ocensm oten lor a emennd	2	1	36.00	36.00			
#	Description	Qty	Net	Total	Amt/Unit																								
1	Slock to see Romix	1	1	260.00	260.00																								
2	Micropla tabocat 1000	1	1	150.00	150.00																								
3	Ocensm oten lor a emennd	2	1	36.00	36.00																								
			<table><tr><td>Subtotal</td><td>4300.00</td></tr><tr><td>Simple tax</td><td>50.00</td></tr><tr><td>Total</td><td>4350.00</td></tr><tr><td>Total</td><td>43,280.00</td></tr></table>			Subtotal	4300.00	Simple tax	50.00	Total	4350.00	Total	43,280.00																
Subtotal	4300.00																												
Simple tax	50.00																												
Total	4350.00																												
Total	43,280.00																												

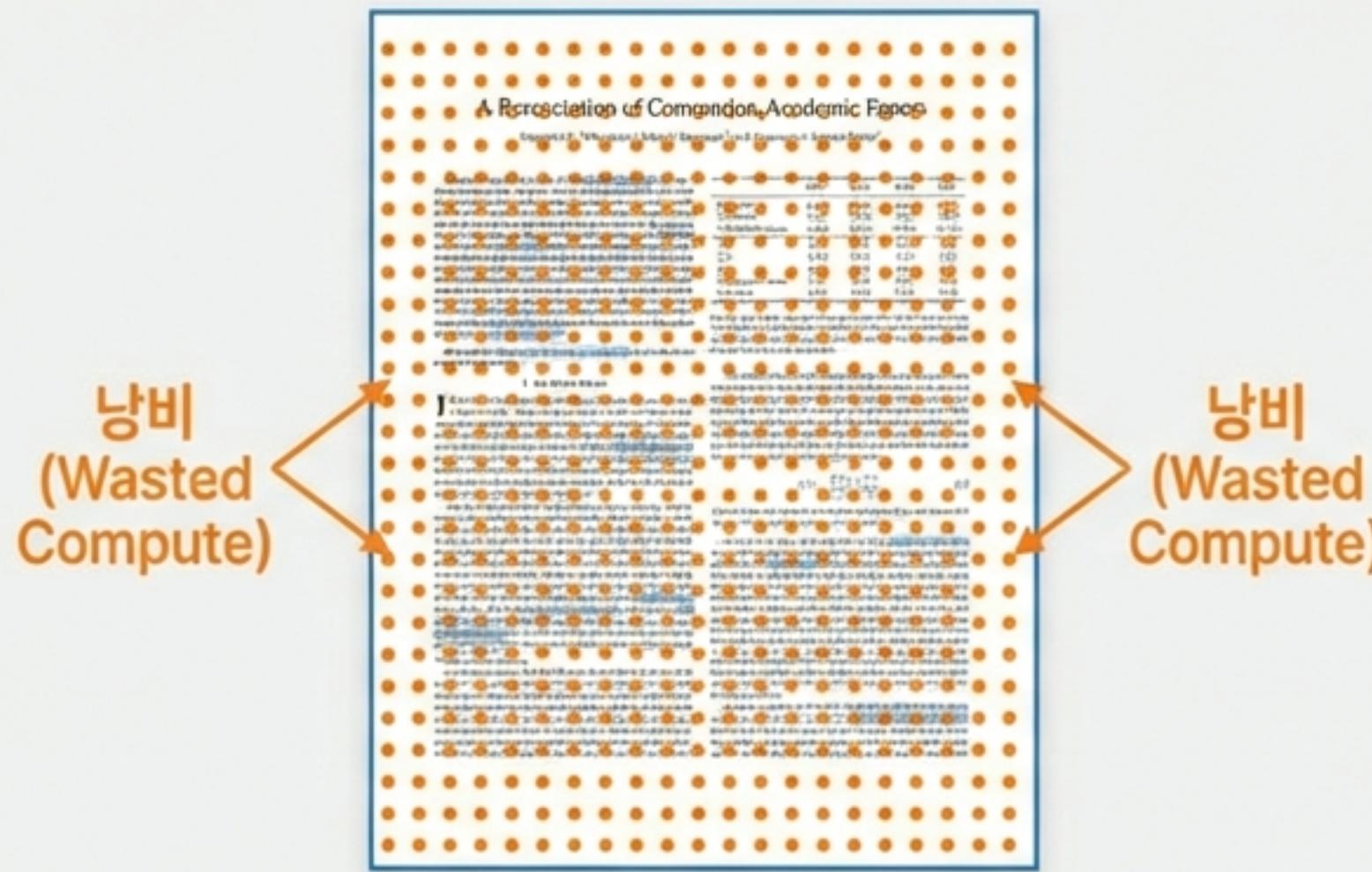
이산적/구조적 정보 (Discrete/Structured Info)

희소한 정보, 행/열 정렬이 핵심.

“풍경을 처리하는 방식으로 표(Table)를 읽으려 하기 때문에 비효율이 발생합니다.”

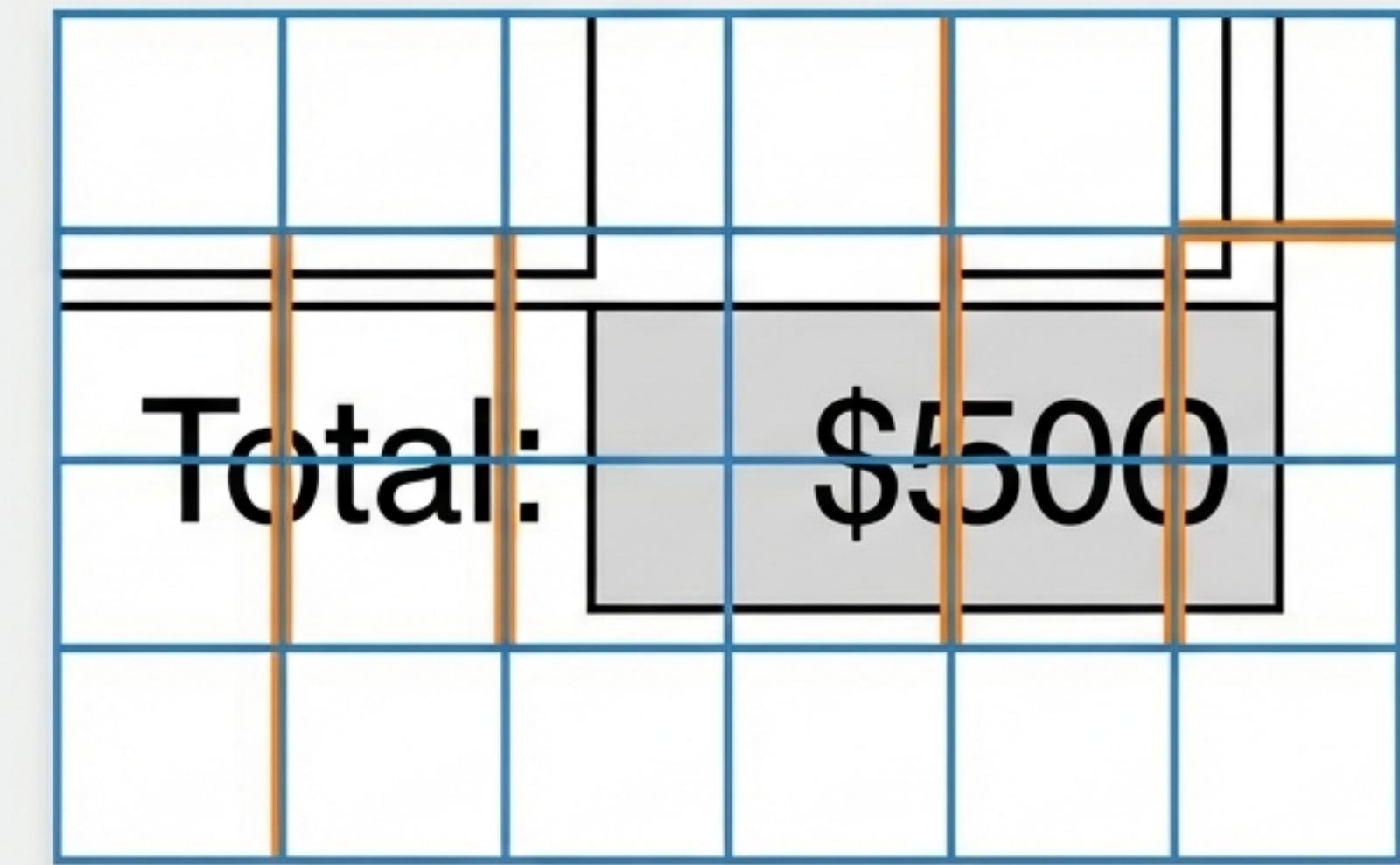
# 획일적 처리의 대가: 비효율성과 환각 현상 (Hallucination)

Inefficiency



불필요한 토큰 할당  
(Inefficient Allocation)

Instability

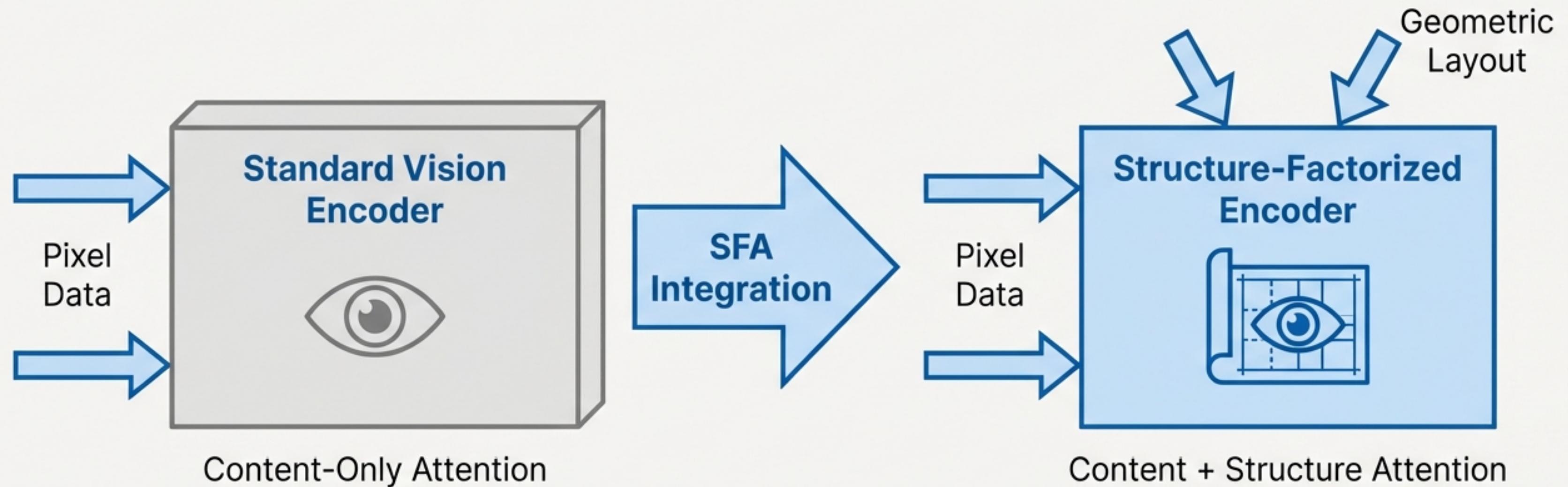


의미론적 구조 파괴  
(Broken Semantic Structure)

텍스트가 없는 여백에 연산 자원을 낭비하고, 행/열 구조를 무시한 패치 분할은 모델의 시각적 그라운딩(Visual Grounding)을 불안정하게 만듭니다.

# 해결책 1: 구조 분해 어텐션 (Structure-Factorized Attention, SFA)

비전 인코더에 '구조적 귀납 편향(Structural Inductive Bias)'을 직접 주입합니다.

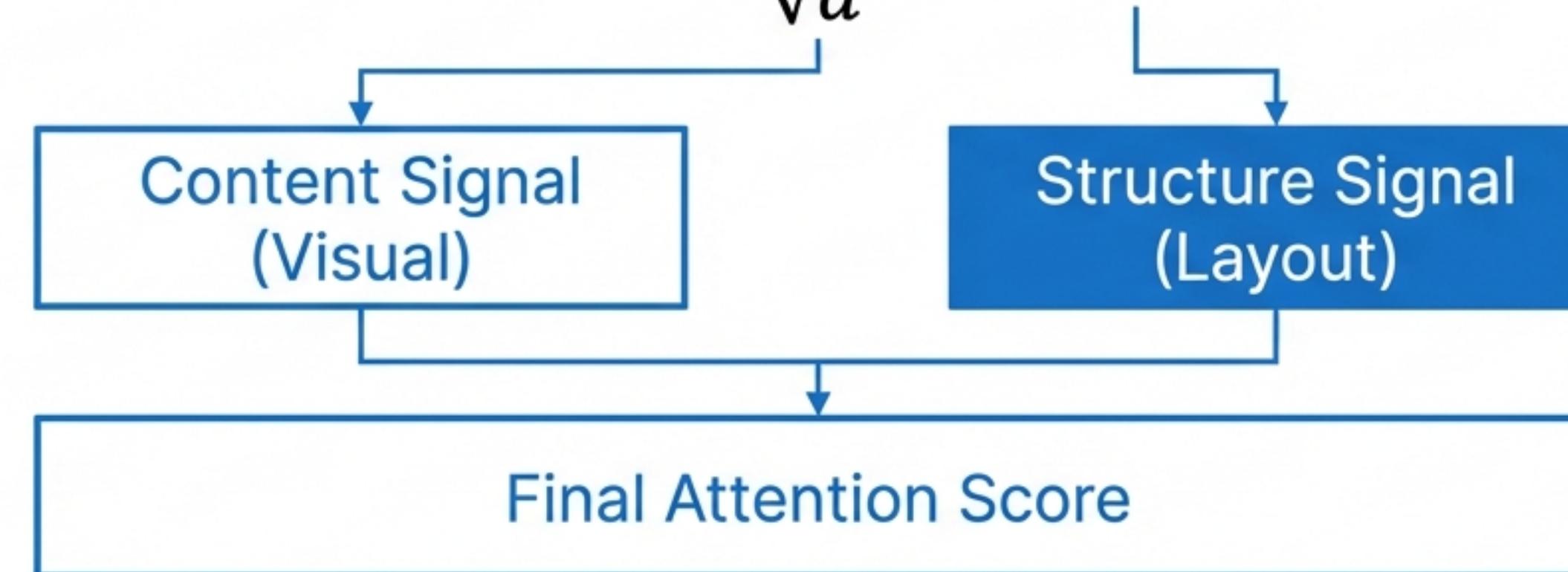


단순히 보는 것(Looking)이 아니라, 레이아웃을 이해하는 것(Understanding Layout).

# 어텐션 커널의 재설계 (Redesigning the Attention Kernel)

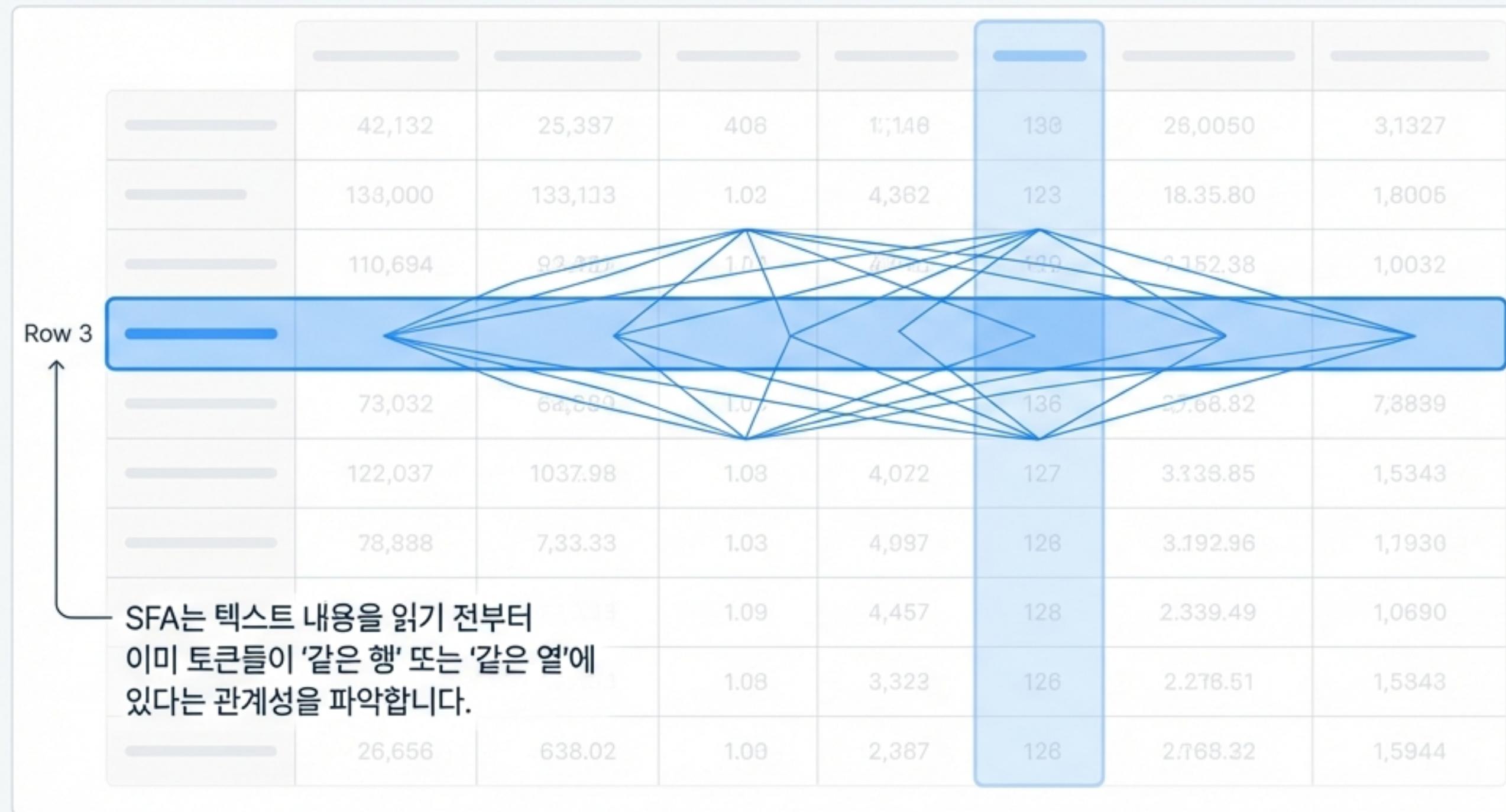
$$\text{Standard: } S_{ij} = \frac{Q_i K_j^\top}{\sqrt{d}}$$

$$\textbf{SFA: } S_{ij} = \frac{Q_i K_j^\top}{\sqrt{d}} + \phi(s_i, s_j)$$



$\phi$ 는 학습 가능한 파라미터(Learnable Parameters)로, 토큰 간의 행(Row), 열(Column), 블록(Block) 관계를 모델링합니다.

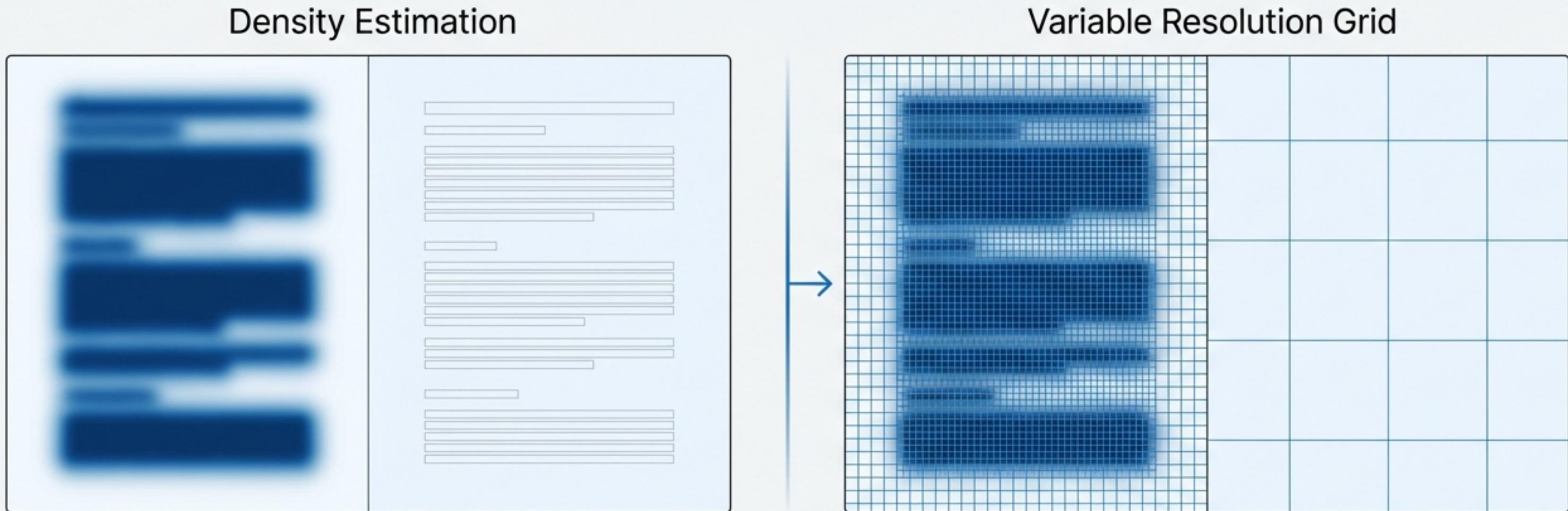
# 행, 열, 그리고 블록 관계의 모델링



**Impact:** 모델 복잡도를 증가시키지 않으면서(Without increasing complexity) 구조적 상호작용을 유도합니다.

# 해결책 2: 적응형 밀도 기반 토큰화 (Adaptive Density-Aware Tokenization)

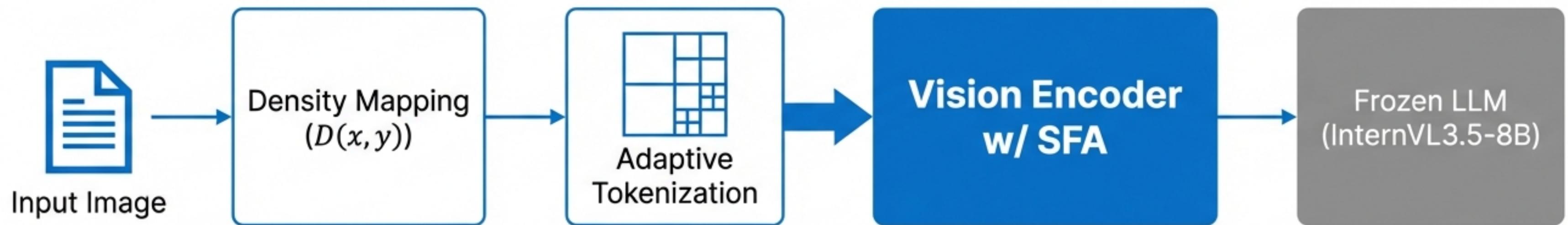
## 예산 재할당 (Budget Reallocation)



고정된 토큰 예산( $N$ ) 내에서 정보 밀도에 따라 해상도를 조절합니다.

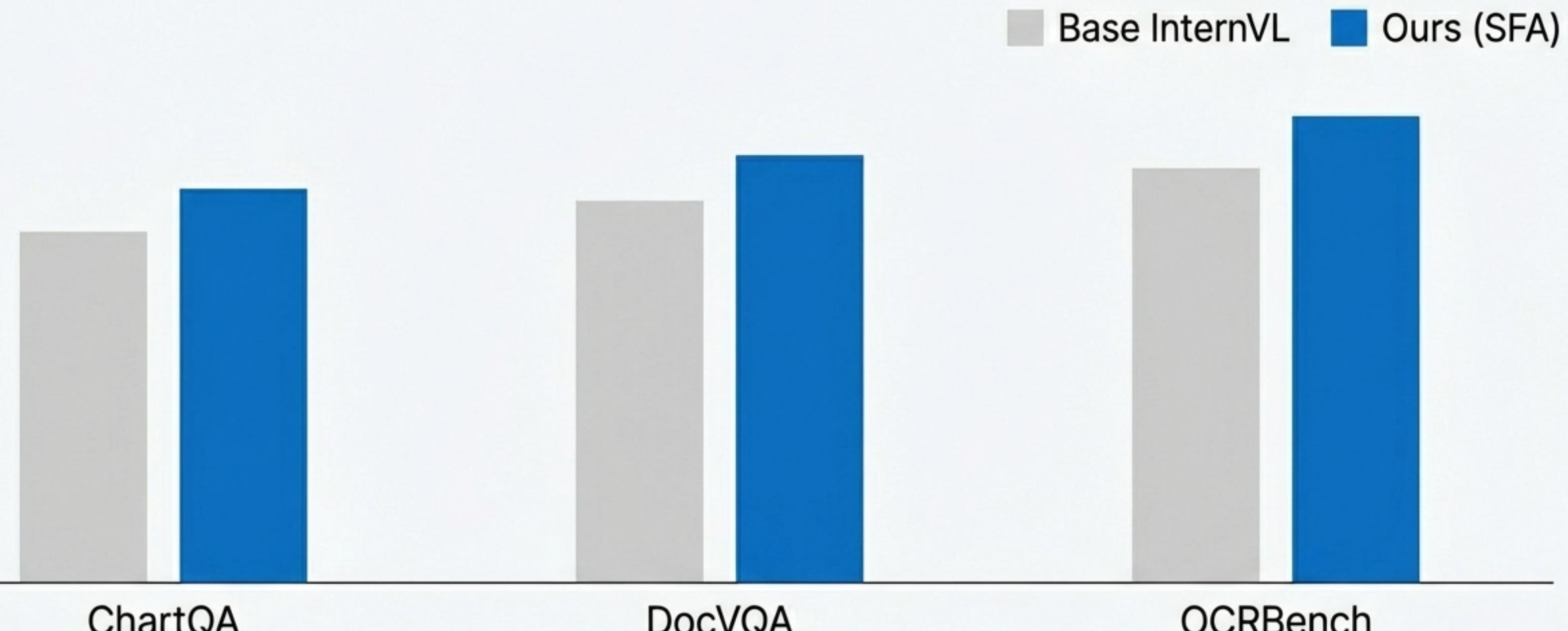
$$\text{Constraint: } \sum_k \frac{\text{Area}_k}{s_k^2} \leq N$$

# 전체 아키텍처 파이프라인 (System Architecture)



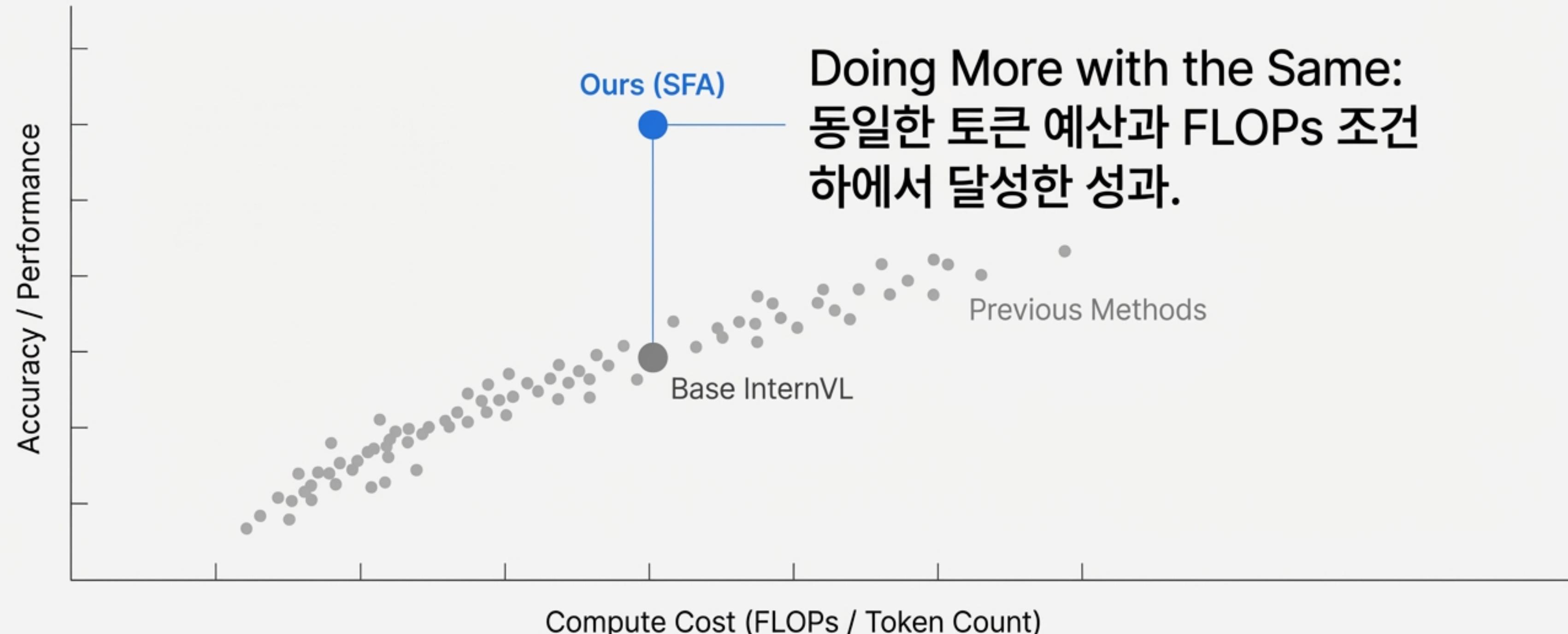
LLM 백본은 동결(Frozen) 상태로 유지하여, 순수하게 인코더 레벨의 개선만으로 성능 향상을 입증했습니다.

# 벤치마크 성능 평가 (Benchmark Performance)



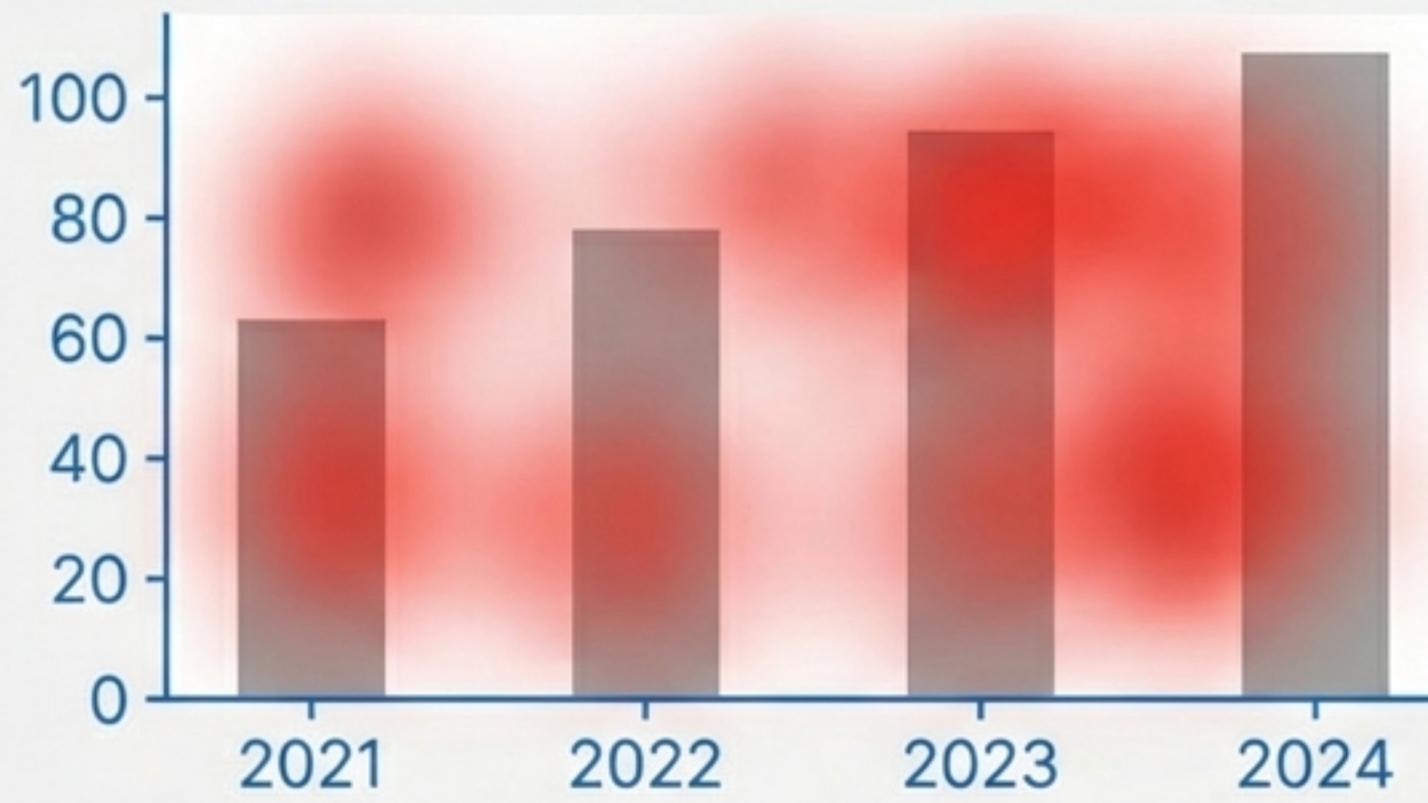
단순히 모델 크기를 키운 것이 아니라, 구조적 이해도를 높여 얻어낸 결과입니다.

# 연산 비용 증가 없는 성능 향상 (Efficiency Analysis)

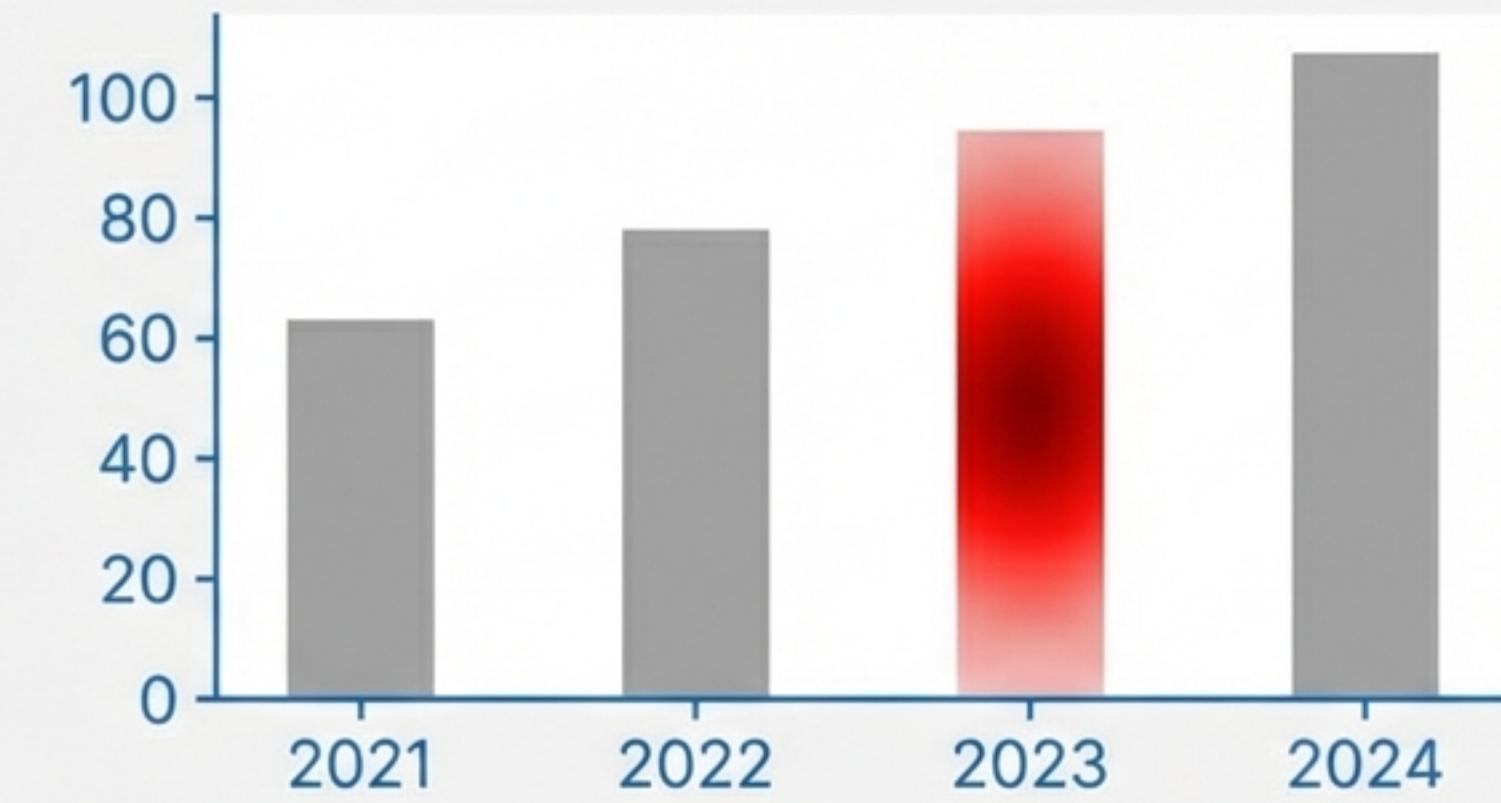


# 시각적 그라운딩(Visual Grounding) 비교 분석

Q: What is the value in 2023?



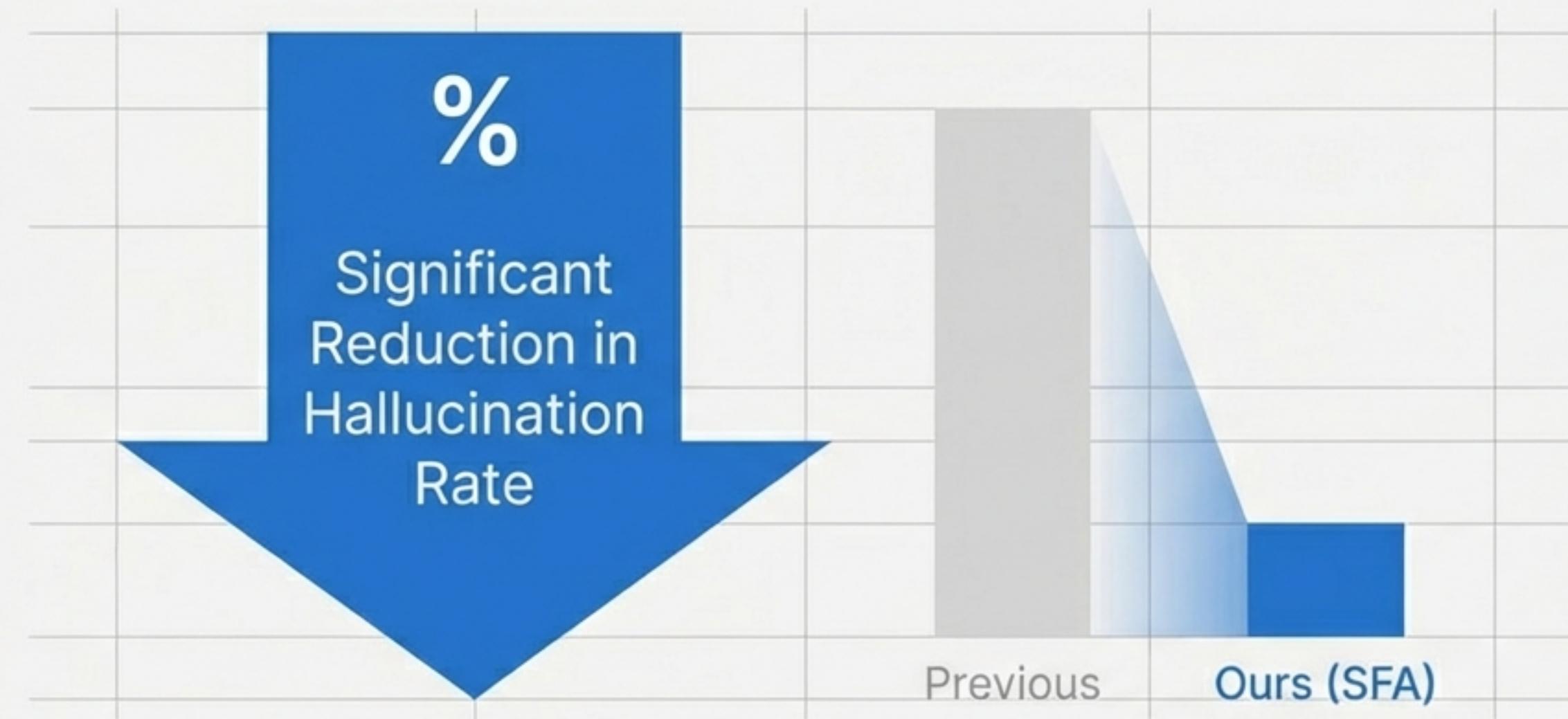
Baseline (Uncertainty)



Ours (Precise Grounding)

SFA는 질문과 관련된 시각적 증거를 정확하게 포착합니다.

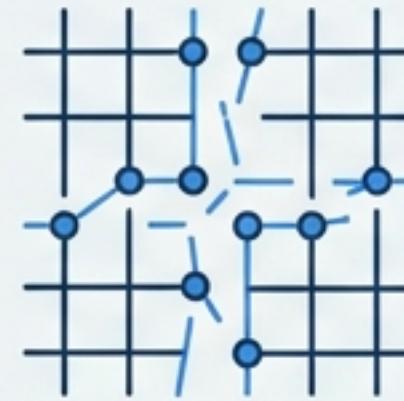
# 구조적 인식을 통한 환각(Hallucination) 감소



문서의 구조를 정확히 파악하면 모델이 없는 정보를 지어내는 비율이 줄어듭니다.

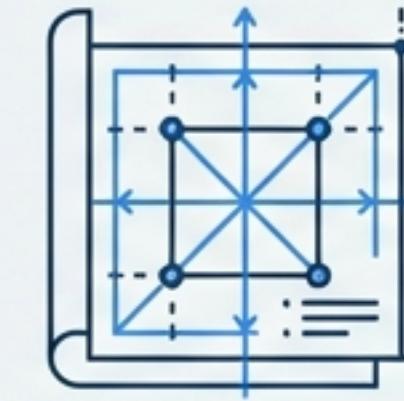
강력한 그라운딩(Strong Grounding)은 곧 신뢰할 수 있는 응답(Reliable Response)으로 이어집니다.

# 결론 및 핵심 요약 (Conclusion & Takeaways)



## The Gap

문서와 자연 이미지 인코더  
간의 구조적 불일치 규명.



## The Innovation

구조 분해 어텐션(SFA)과  
적응형 토큰화를 통한 해결.



## The Result

추가 연산 비용 없이 그라운딩  
안정성과 성능 동시 확보.

인코더 레벨의 구조적 귀납 편향(Structural Inductive Bias)은  
신뢰할 수 있는 문서 이해 모델의 필수 요소입니다.