Nano Banana and Its Prospect in Medical Imaging

Presented by: Abdullah Mamun

Date: 17 September 2025

Email: a.mamun@asu.edu

Original Edited by Nano Banana

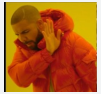
Nano Banana

abdullah-mamun.com

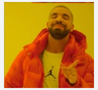
𝕏: @AB9Mamun

Original

Prompt: "Replace the background with the Grand Canyon."



GPT 5



Nano Banana

How is Google marketing Nano Banana?

Here are a few things to try as you explore this new image editing capability:

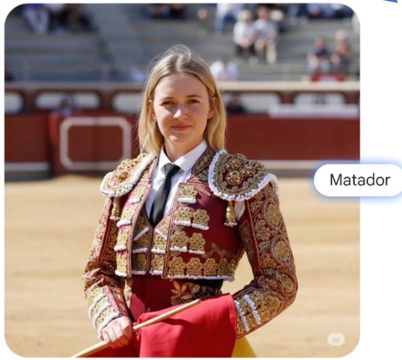
- **Give yourself a costume or location change:** Upload a photo of a person or pet, and our model will keep their look the same in every image as you place them in new scenarios. Try putting yourself in different outfits or professions, or even see how you'd appear in another decade — all while still looking like you.

Example Prompt

Reimagine this person as a matador inside a bullfighting ring.

AI-generated

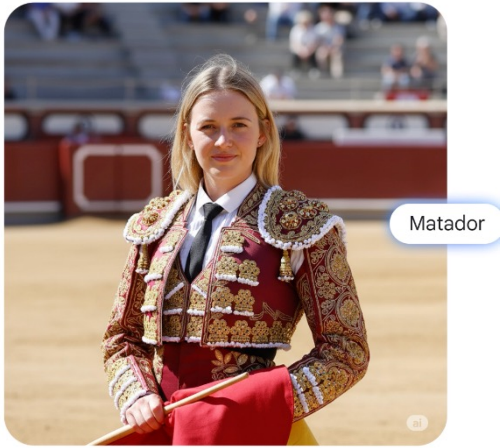
Original



Different prompt

Different prompt

Nano Banana vs GPT 5



Nano Banana

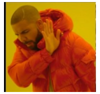
Original

GPT 5 (DALL·E through GPT-4o’s image generation capabilities)

Prompt: Reimagine this person as a matador inside a bullfighting ring.

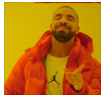
Nano Banana vs GPT 5

This looks nothing like me!



Original

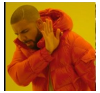
GPT 5
(DALL·E through GPT-4o's
image generation capabilities)



Nano Banana

Prompt: Replace the background with the
Grand Canyon.

Nano Banana vs GPT 5

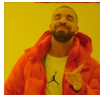


Still not me!

Original

GPT 5 (Latest DALL E with latest GPT version)

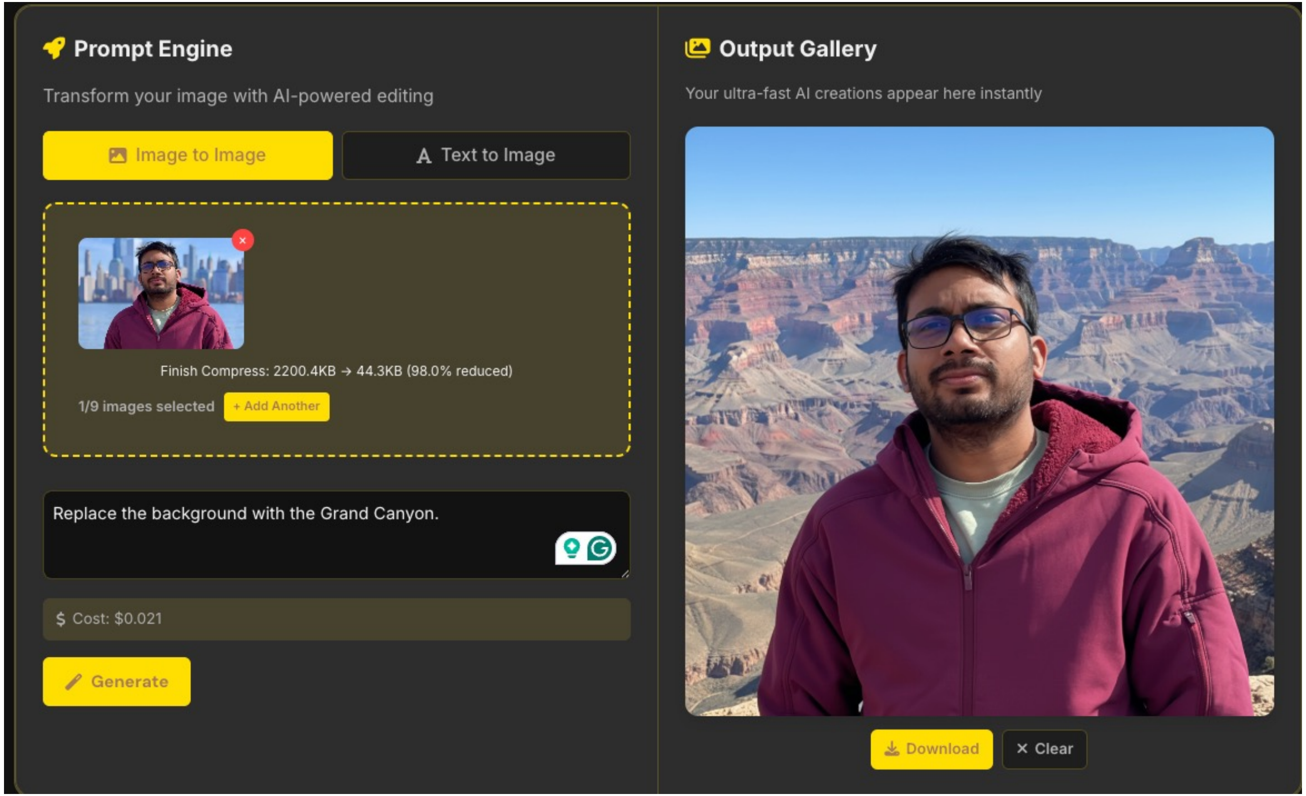
2nd chance



Nano Banana

Prompt: Replace the background with the Grand Canyon.

Nano Banana was not only highly superior in the quality, but also much faster in generating the image.



Effect comparison

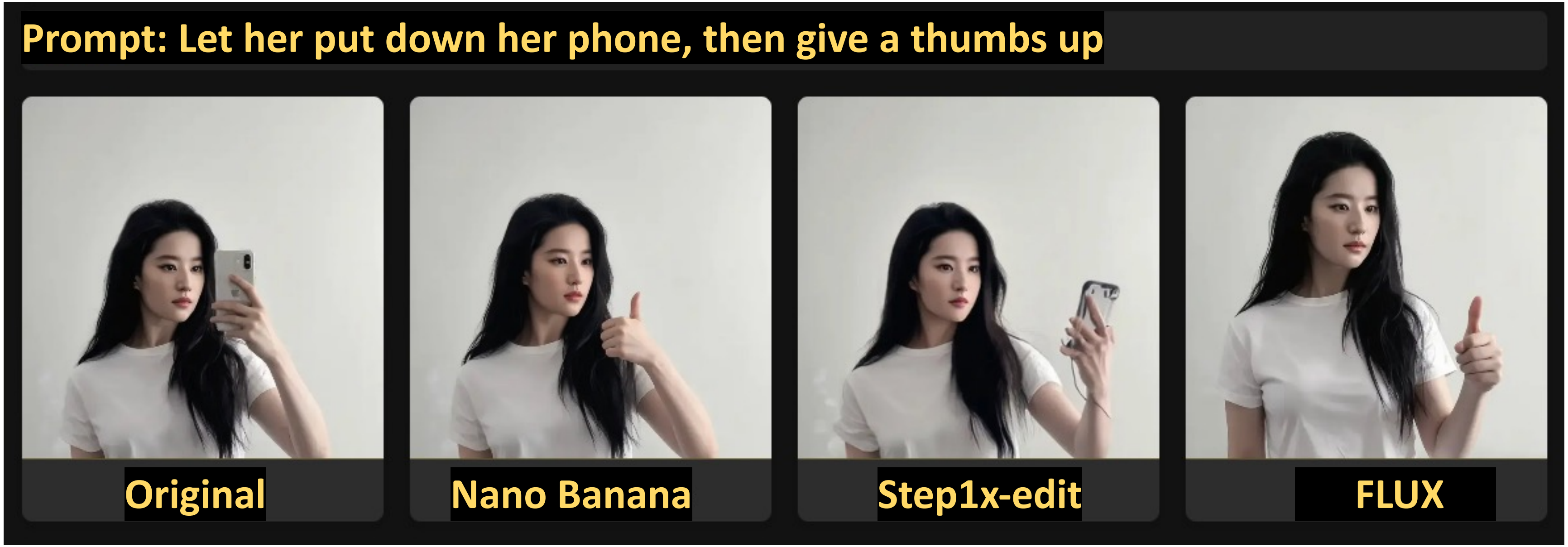
Prompt: Replace the clock in the image with a Rolex

Original | Nano Banana | Gemini 2.0 Flash | GPT-Image

Prompt: Replace David with Iron Man and create it in a magazine cover poster style

Original | Nano Banana | Gemini 2.0 Flash | SEED

Another example



Strengths of Nano Banana

Advanced AI Reasoning

Nano Banana AI can "think" about the context of your prompts and apply reasoning to generate accurate, realistic images accordingly.

3D Object Editing

Advanced neural networks comprehend 3D relationships within 2D images, allowing you to manipulate objects with precision while preserving the rest of the image.

Intelligent Image Generation

Create stunningly realistic images from text descriptions with the Nano Banana model's remarkable accuracy and attention to detail.

Consistency Preservation

Sophisticated algorithms maintain perfect consistency across edits while understanding the overall composition and style.

Deep Prompt Understanding

Proprietary AI architecture enables the Nano Banana model to "think through" image generation tasks with logical reasoning, interpreting exactly what you want.

Context-Aware Editing

Unlike traditional image editors, Nano Banana combines deep learning with reasoning capabilities to understand not just what you want to create, but why and how it should appear.

But how does it work?

Real name of Nano Banana

- Not open-source
- Gemini 2.5 uses mixture of experts architecture like others.

- May consider reading the Gemini 2.5 paper by Google

- **Most likely diffusion model** for image generation/editing (according to the Internet, not confirmed by Google)

- My guess: maybe one key trick is localizing the part of the image that will remain unchanged and generating the other pixels – or maybe different fine tuning with labeled data?

https://arxiv.org/abs/2507.06261

Gemini 2.5 Flash Image (Nano Banana)

	Gemini 1.5 Flash	*Gemini 1.5 Pro*	**Gemini 2.0 Flash-Lite**	**Gemini 2.0 Flash**	**Gemini 2.5 Flash**	**Gemini 2.5 Pro**
Input modalities	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio
Input length	1M	2M	1M	1M	1M	1M
Output modalities	Text	Text	Text	Text, Image*	Text, Audio*	Text, Audio*
Output length	8K	8K	8K	8K	64K	64K
Thinking	No	No	No	Yes*	Dynamic	Dynamic
Supports tool use?	No	No	No	Yes	Yes	Yes
Knowledge cutoff	November 2023	November 2023	June 2024	June 2024	January 2025	January 2025

Table 1 | Comparison of Gemini 2.X model family with Gemini 1.5 Pro and Flash. Tool use refers to the ability of the model to recognize and execute function calls (e.g., to perform web search, complete a math problem, execute code). **currently limited to Experimental or Preview, see Section 2.7. Information accurate as of publication date.*

Why do we care as Healthcare AI researchers?

366v1 [eess.IV] 31 Aug 2025

Can General-Purpose Omnimodels Compete with Specialists? A Case Study in Medical Image Segmentation

Yizhe Zhang *zhangyizhe@njust.edu.cn*
School of Computer Science and Engineering
Nanjing University of Science and Technology

Qiang Chen *chen2qiang@njust.edu.cn*
School of Computer Science and Engineering
Nanjing University of Science and Technology

Tao Zhou *taozhou@njust.edu.cn*
School of Computer Science and Engineering
Nanjing University of Science and Technology

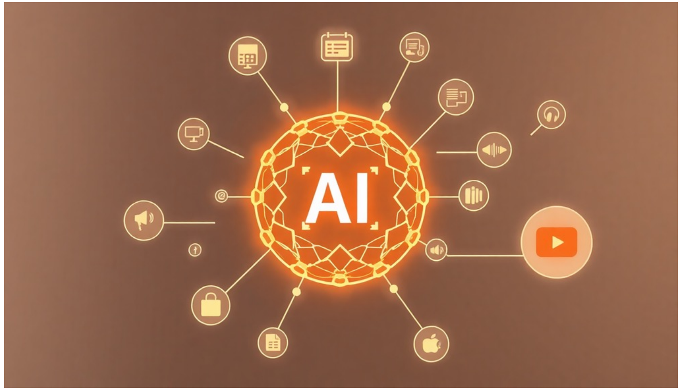
Abstract

The emergence of powerful, general-purpose omnimodels capable of processing diverse data modalities has raised a critical question: can these "jack-of-all-trades" systems perform on par with highly specialized models in knowledge-intensive domains? This work investigates this question within the high-stakes field of medical image segmentation. We conduct a comparative study analyzing the zero-shot performance of a state-of-the-art omnimodel (Gemini 2.5 Pro, the "Nano Banana" model) against domain-specific deep learning models on three distinct tasks: polyp (endoscopy), retinal vessel (fundus), and breast tumor segmentation (ultrasound). Our study focuses on performance at the extremes by curating subsets of the "easiest" and "hardest" cases based on the specialist models' accuracy. Our findings reveal a nuanced and task-dependent landscape. For polyp and breast tumor segmentation, specialist models excel on easy samples, but the omnimodel demonstrates greater robustness

Summary of the paper: Omnimodels vs specialist models

- Compares a state-of-the-art omnimodel (Gemini 2.5 Pro) against domain-specific models.
- Focuses on three distinct tasks: polyp, retinal vessel, and breast tumor segmentation.
- Analyzes performance at the extremes: the "easiest" and "hardest" cases for specialist models.
- Reveals a nuanced, task-dependent landscape of complementary strengths rather than simple superiority.

A Paradigm Shift in AI



The Specialist Era

- For decades, medical AI progress was driven by highly specialized models (e.g., U-Net).
- These models were designed and trained on curated datasets for a single, well-defined task.
- They achieved high performance through focused expertise.

The Generalist Revolution

- The latest evolution is the "omnimodel" (e.g., Gemini).
- Pre-trained on vast, web-scale data.
- Can perform an array of tasks across text, images, and other modalities.
- Demonstrates remarkable zero-shot capabilities.

The Central Research Question

- Can the broad, generalized knowledge of an omnimodel compete with the focused expertise of a specialist?
- Especially in domains where precision, reliability, and safety are paramount.

Omnimodel Promise

Versatility and reduced reliance on task-specific training data.

Specialist Strength

Nuanced perceptual capabilities from architectures optimized for a single task.

The Unknown

Can versatility match the fine-grained performance needed for clinical-grade segmentation?

Experimental Design

Tasks & Datasets

- Polyp Segmentation (CVC-ColonDB)
- Retinal Vessel Segmentation (FIVE)
- Breast Tumor Segmentation (BUSI)

Models

- Specialists: HSNet, U-Net, Mask2Former
- Omnimodel: Gemini 2.5 Pro (Zero-Shot)

Methodology

- Performance Stratification: Top 5% best-performing (Easy) vs. Bottom 5% worst-performing (Hard) samples
- Evaluation Metrics: Dice Coefficient (Overlap) & HD95 (Boundary)

Dice: higher is better
HD95: lower is better

Prompts for Omnimodels

Prompting Strategy. We interacted with Gemini 2.5 Pro via its image generation function using task-specific prompts designed to elicit a binary segmentation mask.

- For polyp segmentation: "Generate a binary segmentation mask of the polyp, ensuring the entire polyp region is fully captured without missing any parts."
- For retinal vessel segmentation: "Generate a binary segmentation mask of the blood vessels, ensuring the vessels are solid (no hollow or broken interiors)."
- For breast tumor segmentation: "Generate a binary segmentation mask of breast ultrasound images, ensuring precise delineation of the lesion region."

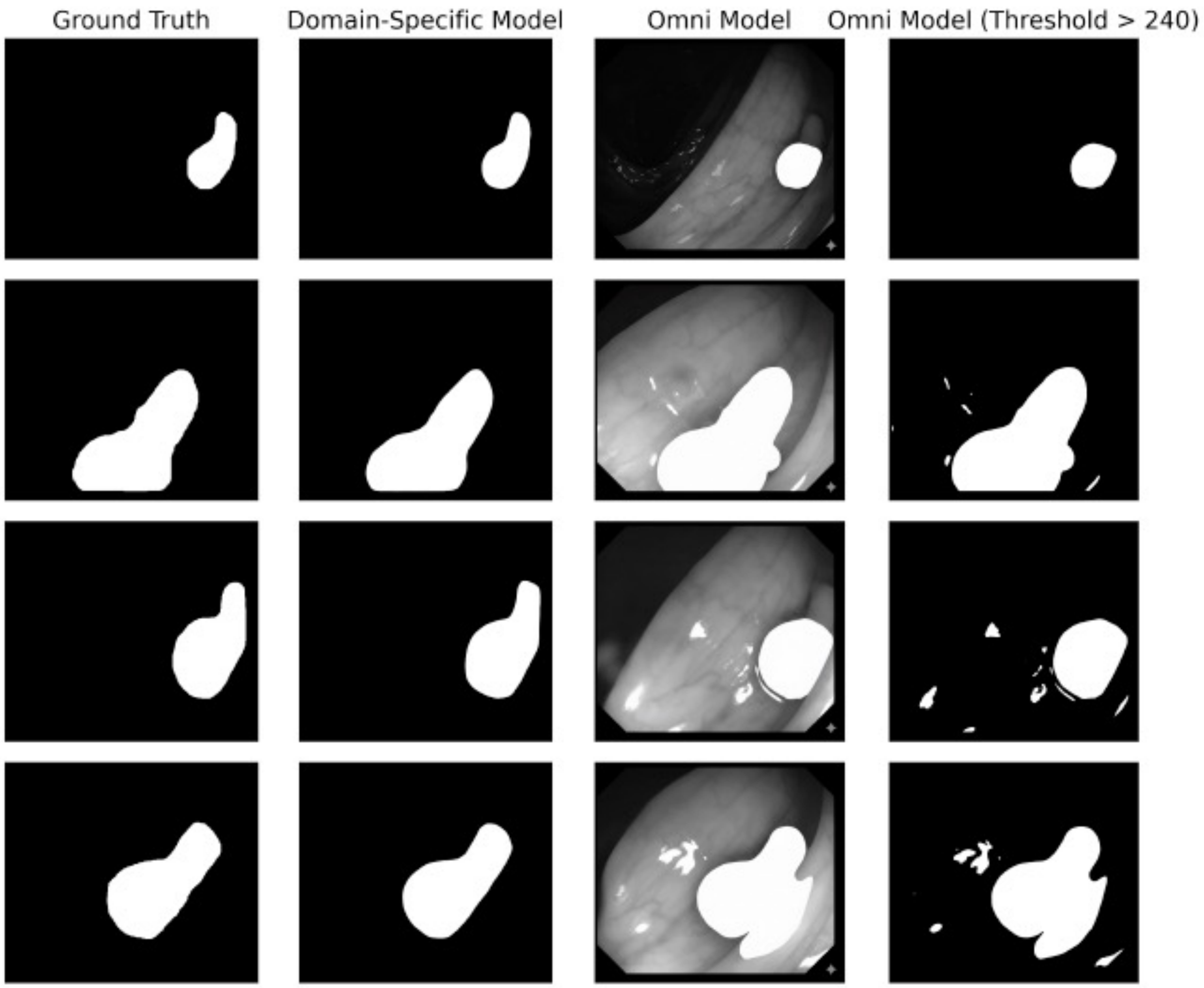
Results: Polyp Segmentation (Colonoscopy)

Table 1: Performance comparison on the best and worst 5% performing samples from the CVC-ColonDB test set. Dice is reported as percentages (↑), while HD95 is in pixels (↓).

Sample Set	Model	Dice (%) ↑	HD95 (pixels) ↓
Easy Samples	Specialist (HSNet)	97.4	6.9
	Omnimodel (Gemini 2.5 Pro)	87.7	87.4
Hard Samples	Specialist (HSNet)	4.3	332.4
	Omnimodel (Gemini 2.5 Pro)	23.6	304.6

For easy samples, specialist model outperformed omnimodel.
For hard samples, omnimodel outperformed specialist model.

Dice: higher is better
HD95: lower is better


Ground Truth
Domain-Specific Model
Omni Model
Omni Model (Threshold > 240)

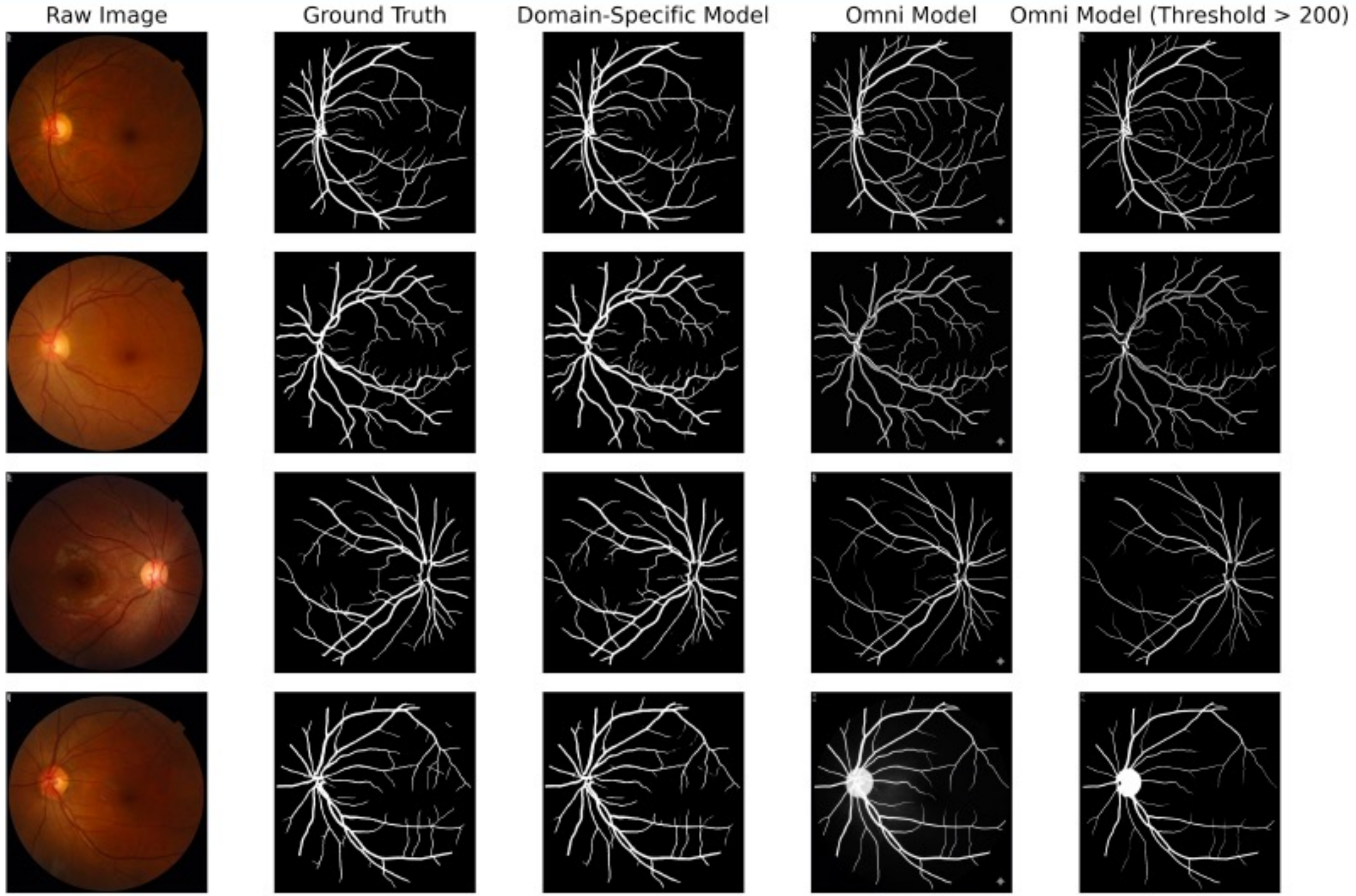
Results: Retinal Vessel Segmentation (Fundus Images)

Table 2: Performance comparison on the best and worst 5% performing samples from the FIVE test set. Dice is reported as percentages (↑), while HD95 is in pixels (↓).

Sample Set	Model	Dice (%) ↑	HD95 (pixels) ↓
Easy Samples	Specialist (U-Net)	91.5	12.0
	Omnimodel (Gemini 2.5 Pro)	62.4	111.0
Hard Samples	Specialist (U-Net)	58.7	174.3
	Omnimodel (Gemini 2.5 Pro)	28.3	408.9

For both easy and hard samples, specialist model outperformed omnimodel.
(see the Key Takeaways slide)

Dice: higher is better
HD95: lower is better



The output of the omnimodel was binarized using a simple threshold and subsequently compared with the ground truth masks for quantitative evaluation.[1]

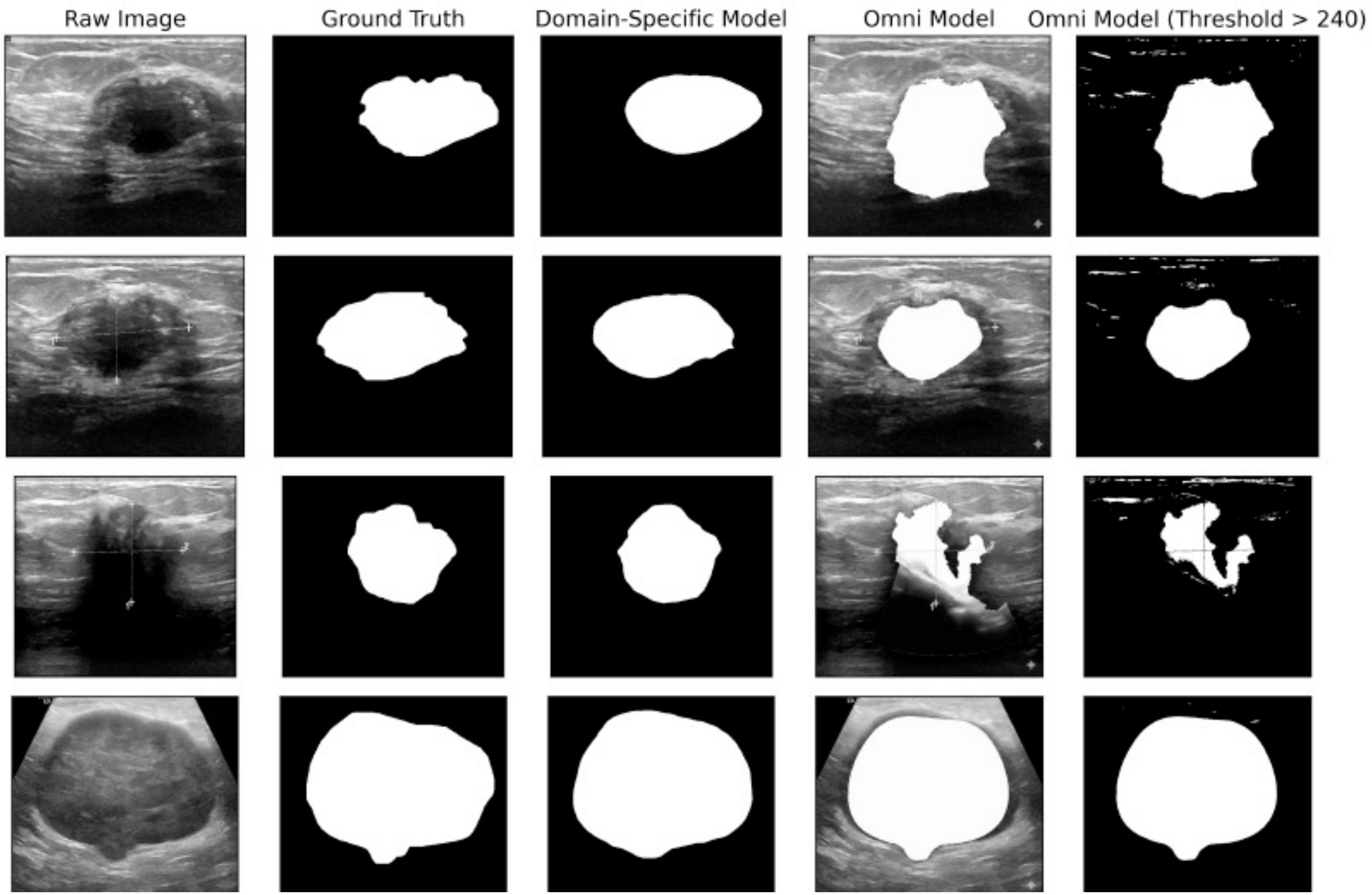
Results: Breast Lesion/Tumor Segmentation (Ultrasound)

Table 3: Performance comparison on the best and worst 5% performing ultrasound segmentation samples. Dice is reported as percentages (↑), while HD95 is in pixels (↓).

Sample Set	Model	Dice (%) ↑	HD95 (pixels) ↓
Easy Samples	Specialist (Mask2Former)	95.2	15.5
	Omnimodel (Gemini 2.5 Pro)	70.5	121.3
Hard Samples	Specialist (Mask2Former)	3.1	263.8
	Omnimodel (Gemini 2.5 Pro)	21.1	198.1

For easy samples, specialist model outperformed omnimodel.
For hard samples, omnimodel outperformed specialist model.

Dice: higher is better
HD95: lower is better



The output of the omnimodel was binarized using a simple threshold and subsequently compared with the ground truth masks for quantitative evaluation.[1]

Key Takeaways

- The relationship is complementary, not direct competition.

- Specialists excel on routine, well-defined cases due to optimized architectures and focused training.

- Omnimodels show robustness on difficult cases, using broad knowledge as a “common sense” safety net when specialist models fail.

- For fine-grained tasks (e.g., retinal vessels), specialized designs remain critical, as omnimodels lack the required precision.

Thank You!

Email: a.mamun@asu.edu

Nano Banana

abdullah-mamun.com

𝕏: @AB9Mamun