The State of Segmentation and Background Removal: A 2025-2026 Technical Monograph

1. Executive Summary: The Great Bifurcation of Computer Vision

The landscape of computer vision, specifically in the domains of segmentation and background removal, has undergone a radical transformation as the industry transitions from late 2025 into 2026. The query for the "best" model no longer yields a singular, monolithic answer. Instead, the field has bifurcated into two distinct, highly specialized directions: the pursuit of **Universal Foundation Models** that prioritize zero-shot generalization across diverse modalities, and the refinement of **Task-Specific Dichotomous Architectures** that achieve hyper-precision in background removal through novel matting strategies.

For the practitioner seeking the optimal solution, the decision matrix now hinges on a critical trade-off between semantic versatility and pixel-perfect fidelity. On one side of the spectrum, Meta's **Segment Anything Model 2 (SAM 2)** has redefined the boundaries of promptable segmentation, introducing streaming memory architectures that unify image and video processing into a cohesive, temporal framework.[1] On the opposite end, the specific task of background removal—often dismissed in previous years as a solved problem—has seen a renaissance with the emergence of **RMBG-2.0**, **BiRefNet**, and **BEN2**. These models have moved beyond simple binary masking to master the complexities of alpha matting, successfully resolving sub-pixel details such as hair strands, transparent glass, and motion blur that historically plagued earlier architectures like U-Net.[4]

Simultaneously, the "Edge Wars" have intensified. The long-standing dominance of the YOLO (You Only Look Once) family is being aggressively challenged by Real-Time Detection Transformers (RT-DETRs). The release of **YOLO11** in September 2025 pushed the efficiency of Convolutional Neural Networks (CNNs) to new heights, yet the emergence of **RF-DETR**—operating under a permissive Apache 2.0 license—has disrupted the status quo by offering superior accuracy-latency trade-offs and eliminating the need for Non-Maximum Suppression (NMS) post-processing.[7]

This comprehensive report serves as a definitive technical guide to this evolving landscape. It dissects the architectures, performance metrics, and deployment realities of the leading models, providing the depth of insight required to make high-stakes engineering decisions in an era defined by rapid algorithmic obsolescence.

2. Theoretical Foundations and Architectural Evolution

To select the "best" model, one must first understand the architectural shifts that define the current generation of computer vision. The progression from 2023 to 2026 has not merely been an exercise in scaling parameters, but a fundamental rethinking of how neural networks perceive and isolate objects.

2.1 The Shift from Semantic to Promptable Segmentation

Historically, segmentation was categorized rigidly: **Semantic Segmentation** classified pixels into predefined classes (e.g., "road," "sky"); **Instance Segmentation** distinguished individual objects (e.g., "car 1," "car 2"); and **Panoptic Segmentation** combined both.[9] These models were trained on closed-sets (like COCO's 80 categories) and failed when presented with novel objects.

The paradigm shift, crystallized by SAM and perfected in **SAM 2**, is the move toward **Promptable Segmentation**. In this regime, the model is class-agnostic. It does not learn "what a car looks like" in a rigid sense, but rather "what an object is" based on geometric and textural coherence. By conditioning the output on user prompts—clicks, bounding boxes, or text—these foundation models achieve zero-shot capability, segmenting objects they have never seen during training, from microscopic cells to extraterrestrial rovers.[2]

2.2 The Matting Renaissance in Background Removal

Background removal is technically distinct from standard segmentation. Standard segmentation outputs a **Binary Mask** (a pixel is either foreground or background, 0 or 1). However, natural images contain "mixed pixels" at boundaries—strands of hair, fur, or smoke where the pixel's color is a linear combination of the foreground and background.

The latest generation of background removal models (RMBG-2.0, BiRefNet) treats the problem as **Dichotomous Image Segmentation (DIS)** combined with **Alpha Matting**. They output a continuous value map (0 to 1, or 0 to 255), representing the *opacity* of the foreground. This evolution is driven by architectural innovations like **Bilateral Reference**, which allows the model to consult high-resolution raw pixel data during the upsampling process to resolve these fine details, rather than relying solely on compressed semantic features.[11]

3. Deep Dive: The Vanguard of Background Removal

For users whose primary requirement is "background removal"—specifically for e-commerce, photo editing, or creative compositing—the general-purpose foundation models are often insufficient due to their lack of matting capability. The SOTA in this domain is currently a three-way contest involving **RMBG-2.0**, **BiRefNet**, and **BEN2**.

3.1 RMBG-2.0: The Industrial Standard

RMBG-2.0, released by Bria AI in late 2024/early 2025, currently stands as the reference standard for commercial-grade background removal. It represents a "data-centric" triumph over purely architectural novelty.[4]

3.1.1 The Data-Centric Advantage

While many open-source models are trained on academic datasets like DIS5K (which can be limited in diversity), RMBG-2.0 utilizes the **BiRefNet architecture** but is trained on a massive, proprietary dataset of over 12,000+ high-quality, pixel-accurate images. This dataset is curated to include challenging commercial scenarios: complex e-commerce products with reflective surfaces, gaming assets, and advertising composites.[4]

- **Legal Safety:** A critical differentiator for enterprise users is the legal provenance of the training data. Bria AI explicitly markets the model as trained on licensed data, mitigating the copyright risks associated with models trained on scraped datasets like LAION.
- **Performance Profile:** Benchmarks indicate that RMBG-2.0 outperforms the original BiRefNet weights by approximately 5% in "usability" scores (90% vs 85% success rate). It is particularly robust in handling "complex backgrounds"—images where the background has high-frequency textures that mimic the foreground, a common failure mode for lesser models.[14]

3.1.2 Deployment and Licensing

The model is released under a **CC BY-NC 4.0** (Creative Commons Attribution-NonCommercial) license. This makes it free for research and personal use but requires a commercial agreement for enterprise deployment. This licensing model reflects the immense value of the proprietary training data.[4]

3.2 BiRefNet: The Architectural Breakthrough

BiRefNet (Bilateral Reference Network) is the architectural backbone upon which RMBG-2.0 is built. Developed by researchers targeting High-Resolution Dichotomous Image Segmentation (HR-DIS), it addresses the fundamental bottleneck of resolution.[5]

3.2.1 The Bilateral Reference Mechanism

Traditional segmentation networks (like U-Net) follow an Encoder-Decoder structure. As the image passes through the encoder, it is downsampled to extract semantic meaning (context). When the decoder upsamples the image to generate the mask, it typically relies on "skip connections" to retrieve spatial details. However, at 4K resolutions, even skip connections lose the finest high-frequency details needed for hair segmentation.

BiRefNet solves this by splitting the task into two cooperative modules:

1. **Localization Module (LM):** Uses a Transformer backbone (typically Swin Transformer) on a downsampled image to understand the *global context* and locate the subject.
2. **Reconstruction Module (RM):** Operates on the *original high-resolution image patches*. It uses the semantic map from the LM as a guide ("look here") but processes the raw pixels to determine the precise boundary.

This **Bilateral Reference**—consulting the semantic map (Inward) and the raw pixels (Outward)—allows BiRefNet to achieve State-of-the-Art (SOTA) results on datasets like DIS5K, HRSOD, and COD, specifically excelling in resolving transparent objects and fine mesh structures.[11]

3.3 BEN2: Confidence-Guided Matting

Released in January 2025, **BEN2 (Background Erase Network 2)** introduces a novel approach known as **Confidence-Guided Matting (CGM)**. This model explicitly acknowledges that segmentation and matting are distinct tasks that require different computational intensities.[6]

3.3.1 The Base-Refiner Architecture

BEN2 operates in two stages:

1. **BEN Base:** A fast, efficient segmentation network that generates a coarse mask and, crucially, a **Confidence Map**. This map identifies pixels where the model is uncertain—typically the transition regions around hair, motion blur, or translucent edges.
2. **BEN Refiner:** A specialized matting network that processes *only* the low-confidence pixels identified by the Base.

3.3.2 Efficiency via Selective Compute

This architecture allows BEN2 to scale gracefully to 4K video. Instead of running an expensive heavy transformer on every pixel of a 4K frame, the Refiner only activates on the <5% of pixels that actually require matting. This results in sharper edges than pure segmentation models while being significantly faster than pure matting models.[12]

- **Licensing:** The Base model is typically MIT licensed (open), while the Refiner or the combined pipeline may have commercial restrictions depending on the specific release repository (Hugging Face vs. proprietary API).[16]

3.4 Summary of Background Removal Landscape

The choice between these three depends largely on the "Build vs. Buy" mentality and the specific edge cases of the user's data.

Feature	RMBG-2.0	BiRefNet	BEN2
Core Philosophy	Data-Centric refinement of SOTA architecture.	Architectural innovation for high-res details.	Two-stage Confidence Guided Matting.
Primary Strength	**Reliability.** Best handling of complex/confusing backgrounds.	**Resolution.** Unmatched detail on hair/fur at 2K/4K.	**Efficiency.** Matting quality at high speed via selective refinement.
License	CC BY-NC 4.0 (Paid Commercial).	MIT (Open Source).	MIT (Base) / Commercial (Refiner).
Output	8-bit Alpha Matte.	Alpha Matte.	Alpha Matte.
Ideal User	Enterprise platforms requiring legal safety.	Researchers, open-source devs needing max quality.	Video processing, developers optimizing for latency.

[4]

---

4. Deep Dive: Universal and Promptable Segmentation

While background removal focuses on the binary separation of foreground and background, **Universal Segmentation** aims to interpret the entire visual scene. This domain is dominated by Foundation Models that offer "zero-shot" capabilities—the ability to segment objects they were never explicitly trained to recognize.

4.1 Segment Anything Model 2 (SAM 2)

Released by Meta in mid-2024, **SAM 2** is the successor to the revolutionary Segment Anything Model. It extends the "promptable" paradigm from static images to the temporal domain of video.[2]

4.1.1 The Architecture of Temporal Continuity

The defining feature of SAM 2 is its **Streaming Memory Architecture**. In video segmentation, the challenge is maintaining the identity of an object as it moves, rotates, or is occluded. SAM 2 solves this by treating video processing as a stream:

- **Memory Bank:** As frames are processed, the model stores feature embeddings of the tracked object in a memory bank. This includes the initial user prompt (e.g., the first click) and recent observations.
- **Memory Attention:** When predicting the mask for the current frame, the model attends to both the current visual input and the stored memories. This allows it to "remember" what the object looked like 100 frames ago.
- **Occlusion Handling:** If an object disappears behind an obstacle, the memory bank retains its features. When the object re-emerges, SAM 2 can re-identify it based on its stored appearance history, recovering the track without user intervention.[3]

4.1.2 Performance and Zero-Shot Capabilities

SAM 2 is trained on the **SA-V dataset**, comprising 51,000 videos and over 600,000 spatial-temporal masklets. This scale grants it unparalleled zero-shot performance.

- **Accuracy:** It provides better boundary accuracy than SAM 1 and operates 6x faster in image-only mode.
- **Interactivity:** In video tasks, it achieves high-quality tracking with 3x fewer interactions (clicks/corrections) than previous SOTA methods.[3]

4.1.3 Limitations for Specific Use Cases

While SAM 2 is the "smartest" model, it is not always the best tool for simple background removal.

- **Output Type:** It generally outputs binary masks, not the subtle alpha mattes produced by RMBG-2.0 or BEN2. This makes it less suitable for compositing hair or transparent objects.
- **Resource Intensity:** Even the "Tiny" (SAM2-t) variant is computationally heavier than specialized CNNs like YOLO, making it challenging for ultra-low-latency edge applications.[18]

4.2 OneFormer: The Multi-Task Unifier

For users who need to understand the *semantics* of a scene—i.e., identifying that a segment is a "car" or a "person" rather than just "object 1"—**OneFormer** remains a critical architecture. Unlike SAM, which is class-agnostic, OneFormer unifies the three core segmentation tasks:

1. **Semantic Segmentation:** Labeling every pixel with a class (e.g., road, sky).
2. **Instance Segmentation:** Identifying individual objects (e.g., pedestrian A, pedestrian B).
3. **Panoptic Segmentation:** A combination of both.

OneFormer uses a **task token** to condition the model. By feeding it the token "the task is semantic segmentation," the model dynamically adjusts its architecture to produce the required output. This versatility is indispensable in robotics and autonomous driving, where the system must simultaneously understand the drivable surface (semantic) and track moving obstacles (instance).[9]

4.3 MultiverSeg: The Medical Specialist

In the specialized domain of clinical research, **MultiverSeg** (MIT, 2025) has emerged to address the scarcity of labeled medical data. Unlike generalist models, MultiverSeg is designed for **interactive, few-shot segmentation** of medical volumes (MRI/CT).

- **Mechanism:** It allows a user to segment a dataset without retraining. By interacting with just a few images (e.g., 9 slices), the model learns the specific anatomical target and propagates this understanding to the rest of the dataset.
- **Impact:** It reduces the user input required by orders of magnitude compared to generic tools, accelerating clinical trials where expert time (radiologists) is the most expensive resource.[19]

---

5. Deep Dive: Real-Time Segmentation at the Edge

For applications where latency is measured in milliseconds—autonomous drones, mobile AR, and industrial inspection—the heavy transformers of SAM 2 are often impractical. The "Real-Time" category is currently a battleground between the established YOLO family and the emerging RT-DETR architectures.

5.1 YOLO11: The Apex of CNN Efficiency

YOLO11, released by Ultralytics in September 2025, represents the culmination of the Convolutional Neural Network lineage. It is designed specifically for speed and ease of deployment.[7]

5.1.1 Architectural Refinements: The C3k2 Block

Building on the success of YOLOv8, YOLO11 introduces the **C3k2 block**, a modification of the Cross-Stage Partial (CSP) network that allows for more variable kernel sizes. This enhances feature extraction capabilities without a proportional increase in parameter count.

- **Efficiency:** YOLO11m (Medium) achieves higher mean Average Precision (mAP) on the COCO dataset than its predecessor YOLOv8m, while using **22% fewer parameters**.
- **Segmentation Head:** The model uses a prototype-based segmentation head. It predicts a set of "mask prototypes" for the entire image and then predicts linear combination coefficients for each detected object. This is highly efficient but can result in masks that are slightly "smoother" and less detailed than pixel-wise classifiers like BiRefNet.[7]

5.2 RF-DETR: The Transformer Disruption

RF-DETR (Real-Time Detection Transformer), championed by Roboflow and released in late 2025, challenges the dogma that Transformers are too slow for the edge. It combines the accuracy benefits of the Transformer architecture with the speed required for real-time systems.[8]

5.2.1 Eliminating the NMS Bottleneck

The Achilles' heel of CNNs like YOLO is Non-Maximum Suppression (NMS). CNNs often detect the same object multiple times with slightly different bounding boxes. NMS is a heuristic post-processing step required to filter these duplicates. It is computationally expensive on CPUs and difficult to optimize on some NPUs.
RF-DETR, being a Transformer, uses set-based prediction. It outputs a fixed set of predictions that are unique by design, completely eliminating the need for NMS.

- **Performance:** The **RF-DETR Seg (Preview)** claims to be **3x faster and more accurate** than the largest YOLO11 model on the COCO Segmentation benchmark. This is a significant breakthrough, suggesting that the "Transformer overhead" has been effectively engineered away.[8]

5.2.2 The Licensing War: Apache 2.0 vs. AGPL

A major factor in the adoption of RF-DETR is its **Apache 2.0 license**.

- **YOLO11 (AGPL-3.0):** The AGPL is a "copyleft" license. If a company uses YOLO11 in a SaaS backend or distributes it in an application, they are generally required to open-source their own software or purchase a paid enterprise license from Ultralytics.
- **RF-DETR (Apache 2.0):** This permissive license allows companies to use, modify, and distribute the model in proprietary commercial products without royalties or open-sourcing their code. This makes RF-DETR highly attractive for startups and corporate R&D teams averse to legal friction.[22]

---

6. Implementation and Deployment: From Cloud to Mobile

The "best" model is ultimately the one that runs effectively on the target hardware. 2025 has seen significant strides in making these complex models portable via ONNX and quantization.

6.1 The ONNX Ecosystem and Web Deployment

The **Open Neural Network Exchange (ONNX)** format has become the universal bridge for deployment.

- **Web-Based Inference:** Models like **RMBG-2.0** and **BiRefNet** are available in ONNX

format, allowing them to run entirely in a user's web browser via onnxruntime-web (using WebAssembly and WebGPU). This eliminates server costs and ensures user privacy, as images never leave the device. However, the high resolution (1024x1024) of these models can strain browser memory limits.[24]
- **BEN2 ONNX Strategy:** BEN2 supports ONNX export for both its Base and Refiner components. This modularity is key for mobile developers; an app might run the lightweight Base model for a real-time camera preview (30 FPS) and only trigger the heavier Refiner model when the user captures the photo, balancing user experience with battery life.[16]

6.2 Mobile Quantization

Deploying SOTA models on Android and iOS requires **Quantization**—reducing the numerical precision of weights from 32-bit floating point (FP32) to 8-bit integers (INT8).

- **BiRefNet-Lite:** Specialized "Lite" versions of BiRefNet are designed to be quantization-friendly. While full 4K processing on a phone is challenging, these lite models (approx. 80-100MB) can deliver "hair-level" precision on mobile NPUs (Neural Processing Units) found in modern Snapdragon and Apple Silicon chips.[25]
- **NPU vs. GPU:** Transformer-based models (like MobileSAM or quantized RF-DETR) increasingly benefit from mobile NPUs, which are optimized for matrix operations, whereas older CNNs like YOLO often run efficiently on the mobile GPU or even the CPU.[28]

---

7. The Commercial Landscape: Models vs. Platforms

The user query often conflates "models" (downloadable weights) with "tools" (SaaS platforms). It is vital to distinguish between the two to understand the market options.

7.1 SaaS Leaders: The "Arena" Champions

Platforms like **Photoroom**, **Remove.bg**, and **Canva** utilize proprietary models that are often heavily finetuned versions of open-source architectures (or entirely custom).

- **The "Arena" Insight:** In blind "Background Removal Arena" tests—where humans vote on the better result between two anonymous models—proprietary commercial tools often rank highly because they include post-processing pipelines (auto-cropping, color correction) that raw models lack. Photoroom, for instance, is consistently ranked #1 in user preference for its ability to handle shadows and product context.[29]
- **Cost vs. Control:** Using an API (like Remove.bg) costs money per image but guarantees a certain Service Level Agreement (SLA). Using a model (like RMBG-2.0) is "free" (compute costs only) but requires engineering effort to host and scale.

7.2 The Role of "Safe AI"

Bria AI (RMBG-2.0) creates a bridge between these worlds. They offer the weights for self-hosting but differentiate through the *license*. By indemnifying users against copyright lawsuits (because their data is licensed, not scraped), they offer the security of a SaaS provider with the technical flexibility of an open-source model. This is increasingly critical for enterprises operating under the EU AI Act.[4]

---

8. Comparative Analysis and Final Recommendations

The selection of the "best" model is not a vertical hierarchy but a horizontal choice based on specific constraints.

8.1 Comparison Matrix

Feature	**RMBG-2.0**	**SAM 2**	**RF-DETR**	**YOLO11**	**BiRefNet**
Primary Task	Background Removal (Matting)	Universal Segmentation (Tracking)	Real-Time Detection/Seg	Real-Time Detection/Seg	HR Background Removal
Output	Alpha Matte (Transparency)	Binary Mask	Binary Mask	Binary Mask	Alpha Matte
Zero-Shot?	Limited (Generic Objects)	**Excellent** (Anything)	No (Closed Set)	No (Closed Set)	Limited
Speed	Moderate (GPU needed)	Slow/Moderate	**Fast (Real-Time)**	**Very Fast**	Moderate
License	CC BY-NC (Commercial Paid)	Apache 2.0	**Apache 2.0**	AGPL-3.0 (Restrictive)	MIT (Open)
Best For...	E-commerce, Marketing	Video Editing, Research	Autonomous Systems, Robotics	Mobile Apps, IoT	Academic Research, Open

					Source

8.2 Scenario-Based Recommendations

Scenario A: The Creative Professional / E-commerce Platform

Requirement: You need to remove backgrounds from product photos, keeping transparent bottles and flyaway hair intact.
Recommendation: RMBG-2.0 or BiRefNet.

- **Why:** These models are the only ones that handle *alpha matting* effectively. SAM 2 will give you a jagged cutout; YOLO11 will be too coarse. RMBG-2.0 is preferred for commercial reliability; BiRefNet is the choice if you need a strictly open-source (MIT) solution and can manage the data provenance risks.

Scenario B: The Video Editor / VFX Artist

Requirement: You need to track a specific person or object through a video clip to apply effects or rotoscoping.
Recommendation: SAM 2.

- **Why:** Its streaming memory architecture renders it superior to any frame-by-frame method. The ability to recover objects after occlusion is a game-changer for video workflows.

Scenario C: The Embedded Engineer (Drones/Robotics)

Requirement: You need to detect and segment cars/pedestrians at 30+ FPS on an edge device (e.g., Jetson, Raspberry Pi).
Recommendation: RF-DETR Seg (or YOLO11).

- **Why:** RF-DETR offers the best accuracy-to-latency ratio in late 2025 and avoids the NMS bottleneck. The Apache 2.0 license is a massive advantage for commercial embedded software. If you are locked into a legacy CNN workflow or need the absolute smallest model footprint (Nano), YOLO11 remains a viable, albeit distinctively licensed, alternative.

Scenario D: The Medical Researcher

Requirement: You need to segment tumors or organs in 3D volumes with minimal labeling effort.
Recommendation: MultiverSeg (or SAM 2).

- **Why:** MultiverSeg is specifically architected for the volumetric and interactive nature of medical imaging, outperforming generalist models in few-shot clinical scenarios.[19]

8.3 Conclusion

As we advance through 2026, the era of the "one model fits all" is over. The "best" model is the one that aligns with your specific bottleneck:

- If your bottleneck is **Quality and Transparency**, choose **RMBG-2.0**.
- If your bottleneck is **Generalization and Tracking**, choose **SAM 2**.
- If your bottleneck is **Latency and Licensing**, choose **RF-DETR**.

The divergence of these technologies indicates a healthy, maturing field where specialization enables performance breakthroughs that a single monolithic model could never achieve.

Comprehensive Analysis filed by: Principal Computer Vision Architect, January 2026.

Works cited

1. What is Segment Anything 2 (SAM 2)? - Roboflow Blog, accessed on January 2, 2026, https://blog.roboflow.com/what-is-segment-anything-2/
2. facebookresearch/segment-anything: The repository provides code for running inference with the SegmentAnything Model (SAM), links for downloading the trained model checkpoints, and example notebooks that show how to use the model. - GitHub, accessed on January 2, 2026, https://github.com/facebookresearch/segment-anything
3. [2408.00714] SAM 2: Segment Anything in Images and Videos - arXiv, accessed on January 2, 2026, https://arxiv.org/abs/2408.00714
4. briaai/RMBG-2.0 - Hugging Face, accessed on January 2, 2026, https://huggingface.co/briaai/RMBG-2.0
5. Introduction to BiRefNet - DebuggerCafe, accessed on January 2, 2026, https://debuggercafe.com/introduction-to-birefnet/
6. [2501.06230] BEN: Using Confidence-Guided Matting for Dichotomous Image Segmentation, accessed on January 2, 2026, https://arxiv.org/abs/2501.06230
7. YOLOv11: An Overview of the Key Architectural Enhancements - arXiv, accessed on January 2, 2026, https://arxiv.org/html/2410.17725v1
8. RF-DETR is a real-time object detection and segmentation model architecture developed by Roboflow, SOTA on COCO and designed for fine-tuning. - GitHub, accessed on January 2, 2026, https://github.com/roboflow/rf-detr
9. Best Image Segmentation Models [2025] - Averroes AI, accessed on January 2, 2026, https://averroes.ai/blog/best-image-segmentation-models
10. The 2025 Guide to Object Detection & Segmentation | by Umar Sadique - Medium, accessed on January 2, 2026, https://medium.com/@umar.sadique/the-2025-guide-to-object-detection-segmentation-d1855133387d
11. Comparison between our proposed BiRefNet and other existing methods for HR segmentation tasks. (a) Common framework[31] - ResearchGate, accessed on January 2, 2026, https://www.researchgate.net/figure/Comparison-between-our-proposed-BiRefNet-and-other-existing-methods-for-HR-segmentation_fig2_386298430
12. BEN: Using Confidence-Guided Matting for Dichotomous Image Segmentation -

arXiv, accessed on January 2, 2026, https://arxiv.org/html/2501.06230v2
13. briaai/RMBG-1.4 - Hugging Face, accessed on January 2, 2026, https://huggingface.co/briaai/RMBG-1.4
14. Bria's New State-of-the-Art Remove Background 2.0 Outperforms the Competition, accessed on January 2, 2026, https://blog.bria.ai/benchmarking-blog/brias-new-state-of-the-art-remove-background-2.0-outperforms-the-competition
15. Bilateral Reference for High-Resolution Dichotomous Image Segmentation - arXiv, accessed on January 2, 2026, https://arxiv.org/html/2401.03407v3
16. PramaLLC/BEN2 - GitHub, accessed on January 2, 2026, https://github.com/PramaLLC/BEN2
17. Naya BEN2 state-of-the-art background removal ComfyUI node - Reddit, accessed on January 2, 2026, https://www.reddit.com/r/comfyui/comments/1if91h4/new_ben2_stateoftheart_background_removal_comfyui/?tl=hi-latn
18. Fast Segment Anything Model (FastSAM) - Ultralytics YOLO Docs, accessed on January 2, 2026, https://docs.ultralytics.com/models/fast-sam/
19. New AI system could accelerate clinical research | MIT News, accessed on January 2, 2026, https://news.mit.edu/2025/new-ai-system-could-accelerate-clinical-research-0925
20. Instance Segmentation - Ultralytics YOLO Docs, accessed on January 2, 2026, https://docs.ultralytics.com/tasks/segment/
21. Comparing YOLO11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment - arXiv, accessed on January 2, 2026, https://arxiv.org/html/2410.19869v2
22. RF-DETR Segmentation Instance Segmentation Model: What is, How to Use - Roboflow, accessed on January 2, 2026, https://roboflow.com/model/rf-detr-segmentation
23. RF-DETR Segmentation Preview: Real-Time, SOTA, Apache 2.0 : r/computervision - Reddit, accessed on January 2, 2026, https://www.reddit.com/r/computervision/comments/1nwl9ub/rfdetr_segmentation_preview_realtime_sota_apache/
24. briaai/RMBG-2.0 at main - Hugging Face, accessed on January 2, 2026, https://huggingface.co/briaai/RMBG-2.0/tree/main/onnx
25. onnx-community/BiRefNet_lite-ONNX - Hugging Face, accessed on January 2, 2026, https://huggingface.co/onnx-community/BiRefNet_lite-ONNX
26. ZygoteCode/RmbgSharp: Enterprise-unique blazing-fast & precise C# .NET 9.0 library for background removal in C# using ONNX Runtime with DirectML. FP32 and FP16 versions! Always open for PRs. - GitHub, accessed on January 2, 2026, https://github.com/ZygoteCode/RmbgSharp/
27. Onnx accuracy · Issue #59 · ZhengPeng7/BiRefNet - GitHub, accessed on January 2, 2026, https://github.com/ZhengPeng7/BiRefNet/issues/59
28. Introducing ONNX Runtime mobile – a reduced size, high performance package for edge devices - Microsoft Open Source, accessed on January 2, 2026,

https://opensource.microsoft.com/blog/2020/10/12/introducing-onnx-runtime-mobile-reduced-size-high-performance-package-edge-devices
29. Background Removal Arena - a Hugging Face Space by bgsys, accessed on January 2, 2026, https://huggingface.co/spaces/bgsys/background-removal-arena
30. Created a background removal arena to benchmark best models : r/StableDiffusion - Reddit, accessed on January 2, 2026, https://www.reddit.com/r/StableDiffusion/comments/1idjztu/created_a_background_removal_arena_to_benchmark/
31. Introducing the RMBG v2.0 Model – The Next Generation in Background Removal from Images - Bria Blog, accessed on January 2, 2026, https://blog.bria.ai/introducing-the-rmbg-v2.0-model-the-next-generation-in-background-removal-from-images