CLASSICAL ENGINEERING VS. DEEP SPATIO-TEMPORAL PARADIGMS IN VIDEO POLYP SEGMENTATION: A SYSTEMATIC COMPARATIVE REVIEW

Badalova Lobar Burhonovna
DIGITAL TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE DEVELOPMENT RESEARCH INSTITUTE
Lobar.fb@gmail.com

Abstract: This systematic review compares three foundational paradigms in Video Polyp Segmentation (VPS) for automated colonoscopy: classical engineering, static 2D deep learning, and recurrent spatio-temporal architectures. While the field has shifted from geometric rules to data-driven networks, balancing frame-level accuracy with temporal consistency remains a critical engineering challenge. Using the multi-center SUN-SEG database, we establish a structural taxonomy by evaluating each paradigm's mathematical formulation, failure modes, and throughput under clinical artifacts like specular reflections, motion blur, and out-of-view events. Our synthesis reveals that classical hand-crafted methods offer deterministic explainability but fail under imaging noise. Memory-less 2D deep networks achieve high spatial accuracy but suffer from boundary flickering and tracking dropouts due to an inter-frame blind spot. Conversely, recurrent spatio-temporal hybrids exhibit superior resilience; by integrating bottleneck gating mechanics, they leverage historical hidden states to stabilize boundaries and project polyp shapes through intense noise without sacrificing real-time throughput. This mapping outlines key architectural trade-offs, serving as a deployment reference for future video-stream intelligence frameworks.

Keywords: *Video Polyp Segmentation; Spatio-Temporal Modeling; ConvLSTM;*

Аннотация: В данном систематическом обзоре сравниваются три основополагающие парадигмы сегментации видеополипов (VPS) для автоматизированной колоноскопии: классическая инженерия, статическое 2D глубокое обучение и рекуррентные пространственно-временные архитектуры.

Хотя область исследований перешла от геометрических правил к сетям, управляемым данными, балансировка точности на уровне кадров с временной согласованностью остается критической инженерной задачей. Используя многоцентровую базу данных SUN-SEG, мы создаем структурную таксономию, оценивая математическую формулировку каждой парадигмы, режимы отказов и пропускную способность при наличии клинических артефактов, таких как зеркальные отражения, размытие движения и события, выходящие за пределы поля зрения. Наш синтез показывает, что классические методы, созданные вручную, обеспечивают детерминированную объяснимость, но терпят неудачу при наличии шума изображения. Глубокие 2D сети без памяти достигают высокой пространственной точности, но страдают от мерцания границ и выпадения отслеживания из-за слепой зоны между кадрами. Напротив, рекуррентные пространственно-временные гибриды демонстрируют превосходную устойчивость; Благодаря интеграции механизмов управления узкими местами, они используют исторические скрытые состояния для стабилизации границ и проецирования форм полипов сквозь интенсивный шум без ущерба для пропускной способности в реальном времени. Это сопоставление описывает ключевые архитектурные компромиссы, служащие ориентиром для развертывания будущих интеллектуальных систем обработки видеопотоков.

Ключевые слова: Сегментация полипов в видео; Пространственно-временное моделирование; ConvLSTM;

Annotatsiya: Ushbu tizimli sharh avtomatlashtirilgan kolonoskopiya uchun Video Polip Segmentatsiyasi (VPS) da uchta asosiy paradigmani taqqoslaydi: klassik muhandislik, statik 2D chuqur o'rganish va takroriy fazoviy-vaqtinchalik arxitekturalar. Ushbu soha geometrik qoidalardan ma'lumotlarga asoslangan tarmoqlarga o'tgan bo'lsa-da, kadr darajasidagi aniqlikni vaqtinchalik izchillik bilan muvozanatlash muhim muhandislik muammosi bo'lib qolmoqda. Ko'p markazli SUN-SEG ma'lumotlar bazasidan foydalanib, biz har bir paradigmaning matematik

formulasini, ishlamay qolish rejimlarini va spekulyar aks ettirishlar, harakat xiralashishi va ko'rinishdan tashqari hodisalar kabi klinik artefaktlar ostida o'tkazuvchanlikni baholash orqali strukturaviy taksonomiyani yaratamiz. Bizning sintezimiz shuni ko'rsatadiki, klassik qo'lda yaratilgan usullar deterministik tushuntirishni taklif qiladi, ammo tasvir shovqini ostida muvaffaqiyatsizlikka uchraydi. Xotirasiz 2D chuqur tarmoqlar yuqori fazoviy aniqlikka erishadi, ammo chegara miltillashidan aziyat chekadi va kadrlararo ko'r nuqta tufayli tushib ketishlarni kuzatadi. Aksincha, takroriy fazoviy-vaqtinchalik gibridlar yuqori chidamlilikni namoyish etadi; to'siq bo'yni darvoza mexanizmlarini birlashtirish orqali ular chegaralarni barqarorlashtirish va real vaqt rejimidagi o'tkazuvchanlikni yo'qotmasdan kuchli shovqin orqali polip shakllarini loyihalash uchun tarixiy yashirin holatlardan foydalanadilar. Ushbu xaritalash kelajakdagi video-oqim razvedka tizimlari uchun joylashtirish ma'lumotnomasi bo'lib xizmat qiluvchi asosiy arxitekturaviy murosalarni bayon qiladi.

Kalit so'zlar: Video polip segmentatsiyasi; Fazoviy-vaqtinchalik modellashtirish; ConvLSTM;

I. Introduction

Automated computer-aided detection and diagnosis (CADe/CADx) systems within continuous video-endoscopy are pivotal in mitigating adenoma miss rates during screening colonoscopies. While early detection of precancerous adenomatous polyps significantly improves patient survival, human diagnostic performance exhibits non-trivial variance driven by procedural fatigue, variable mucosal visualization, and the morphological subtlety of flat or sessile serrated lesions. Consequently, engineering robust Video Polyp Segmentation frameworks has become an imperative frontier in biomedical image computing.

The vision architectures powering gastrointestinal endoscopy have transitioned through three distinct evolutionary phases. Early clinical setups relied on classical, hand-crafted engineering frameworks utilizing rule-based geometric equations and variational boundary propagation schemes, such as active contours. Although

structurally deterministic and computationally lightweight, these formulations exhibit extreme vulnerability to the chaotic environments of live clinical streams. The introduction of deep learning redefined the field, transitioning localization workflows from rigid, human-designed rules to highly expressive, data-driven spatial feature representations.

Despite achieving high pixel-level accuracy, deploying static, memory-less 2D deep networks directly into continuous video streams introduces severe operational failure modes. Because these architectures evaluate incoming video frames as isolated snapshots, they possess an inherent "inter-frame blind spot." In a clinical setting, the visual field is continuously corrupted by dynamic artifacts, including intense specular reflections, fast motion blur from camera shifts, and out-of-view events. When processing these corrupted sequences, static networks exhibit high-frequency boundary flickering and acute tracking dropouts, lacking the architectural mechanisms to leverage context from adjacent temporal frames.

To overcome these limitations, the computer vision community has shifted toward deep recurrent spatio-temporal architectures. By embedding temporal gating mechanics—such as Convolutional Long Short-Term Memory (ConvLSTM) modules—at compressed feature bottlenecks, these architectures maintain a continuous spacetime representation of the target sequence. This temporal memory acts as a predictive filter that stabilizes boundary fluctuations and projects estimated polyp shapes through intense, high-frequency noise fields, achieving high tracking resilience while maintaining the real-time throughput (>30 FPS) required for intraoperative surgical deployment.

Contributions of this Review

This systematic review addresses a critical gap in the literature by providing an authoritative, mathematically grounded cross-examination of the structural paradigms governing video-stream intelligence in endoscopy. The specific contributions of this work are three-fold:

1. Structural Taxonomy Mapping: We establish a formal mathematical and structural taxonomy categorizing automated lesion localization into deterministic classical engineering, memory-less 2D deep learning, and sequential recurrent spatio-temporal hybrids.

2. Dynamic Artifact Benchmarking: We evaluate and cross-examine the specific failure modes of each paradigm under real-world video-endoscopic artifacts (specular reflection, motion blur, and out-of-view tracking re-entry) benchmarked against standardized performance profiles from the multi-center SUN-SEG database.

3. Clinical Deployment Reference: We provide an engineering analysis of the trade-offs between mathematical explainability, resource footprint, computational throughput, and tracking stability, serving as an authoritative reference for real-time deployment strategies in next-generation surgical suites.

II. Comparative Analysis of Endoscopic Polysegmentation Taxonomies

Paradigm 1: Classical Engineering Frameworks (Deterministic Video Processing)

A. Operational Mechanics & Functional Principles

Classical engineering frameworks rely strictly on hand-crafted mathematical formulations and rule-based heuristics to isolate anomalies by leveraging pixel-level color distribution, local texture anomalies, and boundary geometry.

• Spatial Feature Extraction: To isolate the boundaries of a polyp from the surrounding mucosal wall, these frameworks exploit the green channel of the standard RGB color space, as it yields the highest signal-to-noise ratio and structural contrast for vascular topology.

• Boundary Localization: Edge localization is achieved via deterministic spatial operators (e.g., Sobel or Canny kernels) or variational boundary propagation models, primarily the Active Contour Model (Snakes).

The optimization process of the Snake Model relies on a continuous balance between intrinsic structural regularization and extrinsic data-driven attraction. The

operational behavior of this variational framework can be decoupled into two primary mechanics.

The internal regularizing force dictates the geometric behavior of the parametric curve independently of the underlying image data, preventing the contour from collapsing or deforming erratically. This is controlled first by the elasticity parameter (α), which regulates the internal tension of the contour. It introduces a continuous contracting force that pulls the parametric curve inward, preventing unbounded expansion and ensuring a tight, compact tracking loop. Working in tandem with elasticity, the stiffness parameter (β) governs the bending resistance of the contour. By penalizing sharp variations in the derivatives along the curve, it suppresses high-frequency geometric noise and forces the boundary to remain smooth, maintaining a topology that mirrors the natural curvature of organic tissue.

The external attraction force establishes the mathematical link between the moving contour and the physical properties of the endoscopic video frame. Edge detection operations within the isolated green channel yield high spatial gradient magnitudes at structural transitions. Because the global optimization framework is structured to minimize the total energy functional, the integration of a negative scalar mapping transforms these peak gradient zones into local energy minima. Mathematically, this negative formulation forces the sharpest visual boundaries of the polyp to act as gravitational attractors, pulling the propagating contour toward the true physical interface of the lesion.

During execution, the parametric contour deforms iteratively across the spatial image coordinates. The curve ceases propagation at the exact operational threshold where the internal regularizing forces reach a state of mathematical equilibrium with the external data attraction of the localized image gradients. This steady state locks the contour coordinates precisely onto the true physical boundary of the adenoma, providing a stable geometric output.

-

The Snake Model Energy-Minimization Equation

$$E_{snake} = \int_0^1 \left(\frac{1}{2} [\alpha |v'(s)|^2 + \beta |v''(s)|^2] + E_{ext}(v(s)) \right) ds \quad (1)$$

Where:

α controls contour elasticity (stretching resistance).

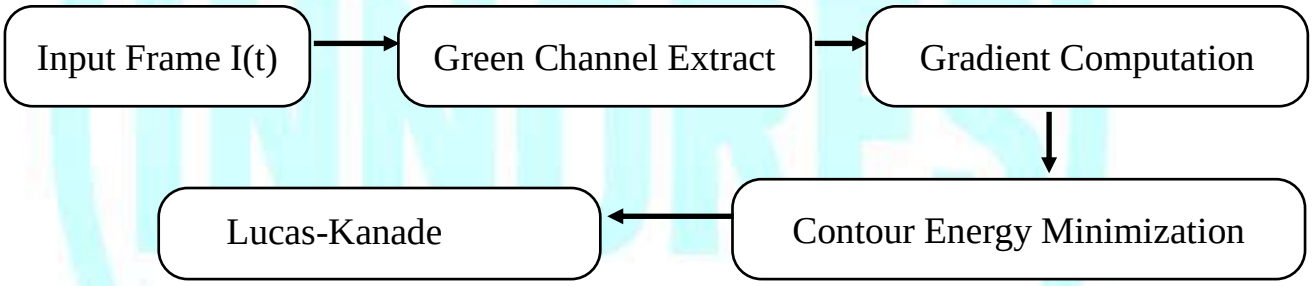
β regulates stiffness (bending resistance).

E_{ext} is derived directly from the image spatial gradient intensity, pulling the contour toward structural boundaries.

Temporal Tracking: Inter-frame dependencies are computed via non-deep tracking algorithms like the Lucas-Kanade Optical Flow equation or Linear Kalman Filtering. These systems assume pixel intensity constancy across short time intervals:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (2)$$

This allows the system to solve a localized first-order Taylor series approximation to generate a 2D velocity vector field, shifting the boundary coordinates from Frame t-1 to Frame t.



B. Technical Advantages and Operational Merits

Classical engineering frameworks exhibit distinct operational merits within specific clinical deployment constraints. First, they possess an ultra-low computational footprint, executing efficiently on legacy, hospital-grade central processing units (CPUs) without requiring high-throughput, dedicated graphics processing units (GPUs). Second, these frameworks offer complete deterministic explainability. Because the system lacks abstract, "black-box" latent layers, every boundary modification and tracking state can be traced directly back to explicit, auditable mathematical thresholds and variational equations. Finally, they feature zero dataset dependency, operating instantaneously without requiring extensive pre-

training data pools, thereby completely eliminating the need for manual, pixel-level ground truth mask annotations.

C. Structural Failure Modes and Vulnerabilities

Despite their determinism, classical methods exhibit severe structural vulnerabilities when subjected to live endoscopic conditions. They generate a characteristically high rate of false positives because non-pathological anatomical elements—such as normal mucosal folds, pooled gastric fluids, fecal residue streams, and background illumination shadows—frequently mirror the exact geometric curvature or color distribution profiles of target adenomas. Furthermore, these systems demonstrate extreme sensitivity to transient video-endoscopic artifacts. Bright specular highlights shift regional pixel intensities to absolute saturation ($Value = 255$), which zeroes out the localized spatial gradient vectors. This mathematical erasure disrupts the active contour tracking loops, causing the parametric curve to collapse.

D. Operational Limitations

The primary operational limitation of these hand-crafted frameworks resides in their poor structural generalization. The rule-based equations are inherently rigid and dependent on sharp contrast boundaries. Consequently, if an endoscope encounters a flat, sessile serrated lesion (SSL) that lacks a distinct vertical shadow profile or a high-magnitude edge gradient, the mathematical heuristics fail to register the anomaly, culminating in a critical diagnostic miss.

Paradigm 2: Static Deep Learning (Memory-Less 2D Frame-by-Frame Networks)

A. Operational Mechanics & Functional Principles

This paradigm replaces hand-crafted geometric rules with deep neural networks that automatically extract multi-scale feature hierarchies from data.

Architectural Flow: Networks utilize an asymmetric encoder-decoder design. Fully Convolutional Networks (FCNs) like U-Net or UNet++ extract low-level textures using an encoder, then map them back to full-resolution binary masks using

a decoder. Advanced architectures like PraNet refine this process by incorporating a Parallel Reverse Attention (RA) module. The RA loop takes deep global features, converts them into a rough area map, and then subtracts that map from higher-level features to force the decoder to focus exclusively on missing boundary details.

Temporal Disconnection: When deployed on live video streams, this framework strips away all temporal information. The continuous 5D video feed is flattened along the timeline into an unlinked 4D batch array:

$$X_{flattened} \in \mathbb{R}^{(B \cdot T) \times C \times H \times W} \quad (3)$$

The model evaluates each incoming frame as an isolated image snapshot,

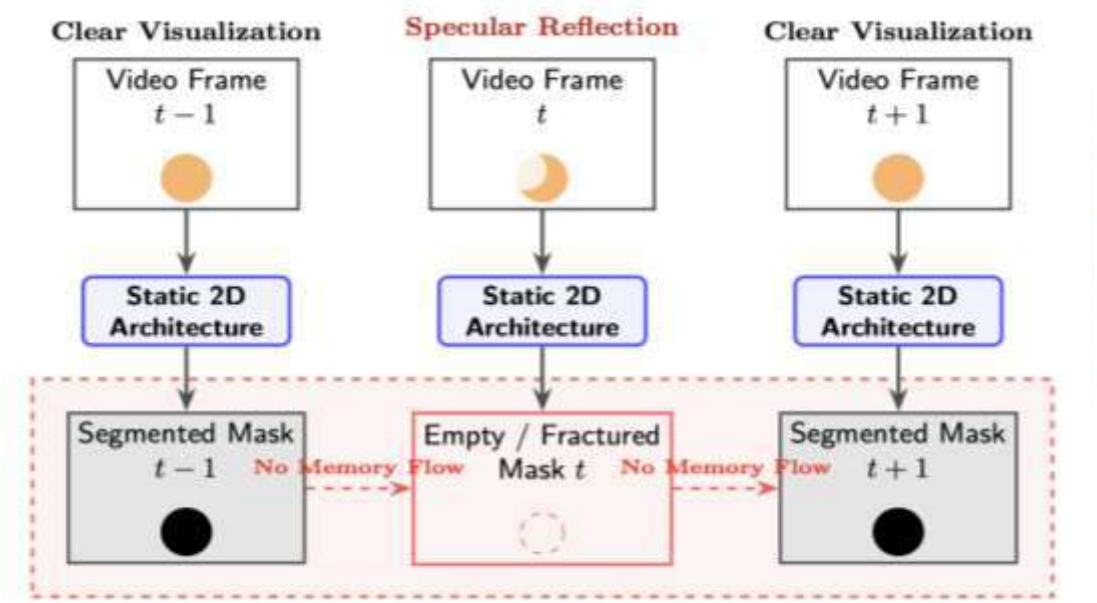


Figure 1. Temporal discontinuity and tracking failure modes in frame-by-frame 2D polyp segmentation pipelines. The illustration highlights the system's inability to leverage historical feature representations from frame $t-1$to preserve boundary consistency through unconstrained imaging noise at frame t.

completely resetting its network activations between steps.

B. Technical Benefits & Advantages

Superior Spatial Precision: Demonstrates high pixel-level accuracy and Dice coefficients when processing high-contrast, well-lit adenomas.

Resilience to Texture Variations: Automated deep kernels adapt smoothly to various polyp classifications, identifying subtle tissue anomalies that hand-crafted classical models miss.

C. Drawbacks & System Failure Modes

High-Frequency Boundary Flickering: Because the architecture operates without memory, minor lighting changes or small camera movements between Frame $t-1$ and Frame t can cause the model's confidence scores to oscillate wildly. This results in an unstable, flickering segmentation mask.

Receptive Field Latency: Processing frames with heavy self-attention mechanisms (such as Vision Transformers or large hybrid models) requires immense computational power, which often drops execution speeds below the 30 frames per second (FPS) standard needed for live intra-operative surgical use.

D. Operational Limitations

The Inter-Frame Blind Spot: These networks are completely blind to temporal context. If a single frame is blurred by sudden movement, a static network cannot leverage the clear, structured features from immediately preceding frames to preserve the tracking mask.

Paradigm 3: Recurrent Spatio-Temporal Networks (Deep Sequential Modeling)

A. Operational Mechanics & Functional Principles

This paradigm treats video polyp segmentation as a continuous spacetime tracking task. Incoming data is kept in its native 5D sequence format:

$$X_{video} \in \mathbb{R}^{B \times T \times C \times H \times W} \quad (4)$$

Where T represents the frame packet length.

The Recurrent Core: Rather than treating frames as isolated events, spatial feature maps pass through a Convolutional Long Short-Term Memory block embedded at the network's compressed feature bottleneck.

Mathematical Word-Equations: Copy this entire structural block into Word. Highlight each line and use the Alt + = command to register them natively within Word's mathematical toolset:

$$i_t = \sigma(W_{xi} * F_t + W_{hi} * H_{t-1} + b_i) \quad \text{[Input Gate]}$$

$$f_t = \sigma\left(W_{xf} * F_t + W_{hf} * H_{t-1} + b_f\right) \quad \text{[Forget Gate]}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot tanh(W_{xc} * F_t + W_{hc} * H_{t-1} + b_c) \quad \text{[Cell State]}$$

$$o_t = \sigma(W_{xo} * F_t + W_{ho} * H_{t-1} + b_o) \quad \text{[Output Gate]}$$

$$H_t = o_t \odot tanh(C_t) \quad \text{[Hidden Output]}$$

Where F_t represents the current incoming spatial feature map, C_t tracks long-term sequence history, and H_t represents the updated spatio-temporal features passed to a progressive 2D upsampling decoder to yield a stabilized output track.

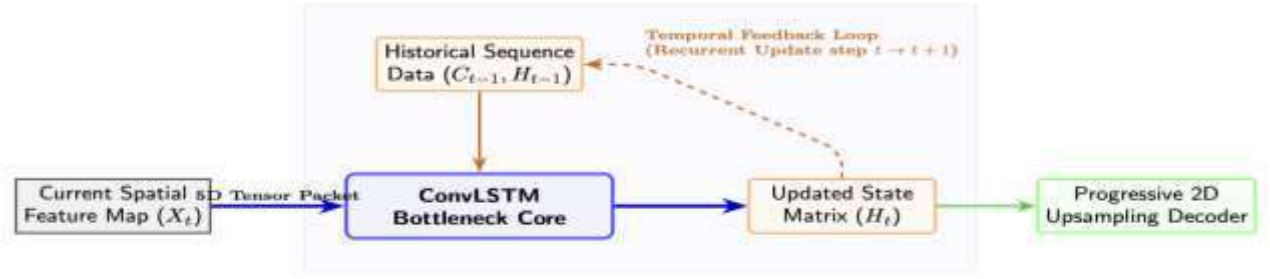


Fig. 2. Spatio-temporal data routing within the ConvLSTM bottleneck (Paradigm 3). The 2D feature map $X_t$and historical states (C_{t-1}, H_{t-1}) are processed via localized convolutions to isolate the updated hidden state H_t. This architecture stabilizes features against endoscopic artifacts before progressive decoding and registers the temporal feedback loop for step $t + 1$.

The ConvLSTM bottleneck integrates temporal memory with spatial awareness by substituting standard matrix multiplications with 2D local convolutions. It preserves spatial dimensions $(H \times W)$ while tracking structures across time through a localized four-step gating process:

B. Technical Benefits & Advantages

Suppression of Mask Flickering: The internal forget gates (f_t) function as an automated temporal filter, smoothing out sudden variations across frames to deliver a stable tracking mask.

Motion Contrast Extraction: The network tracks how an object moves relative to its background over time. This enables it to isolate flat, sessile adenomas that share the same color and texture as the surrounding colon wall based purely on their distinct motion profiles during camera shifts.

Bottleneck Real-Time Performance: By placing the recurrent blocks exclusively at the highly compressed bottleneck layer rather than processing full-resolution maps, networks minimize computational overhead to achieve processing speeds exceeding 110 FPS on clinical graphics workstations.

C. Drawbacks & Failure Modes

Compounding Error Accumulation: If the network generates inaccurate mask predictions over several consecutive frames, these errors pollute the long-term hidden memory state (C_t), leading to persistent tracking lag in subsequent frames.

Complex Optimization Pipelines: Training requires specialized 5D data loaders and multi-frame loss functions, such as Video BCE-Dice-Boundary loss, which significantly increases training times and complexity.

D. Operational Limitations

Vanishing Gradients Across Extended Sequences: Standard ConvLSTMs are designed to process shorter sequence bundles (T = 5 to T = 10). If a polyp remains obscured by intense noise, bubbles, or water jet washouts over an extended interval (e.g., more than 60 frames), the tracking history can degrade, requiring an explicit system reset.

III. CONCLUSION

The technological evolution of video-based polyp segmentation within automated colonoscopy frameworks can be categorized into three distinct operational paradigms. Each paradigm features a fundamental divergence in mathematical

formulation, tensor dimensionality, and structural robustness when subjected to live clinical artifacts.

1. Paradigm 1: Classical Engineering Frameworks

Classical frameworks rely entirely on deterministic, hand-crafted heuristics. By processing inputs as isolated two-dimensional spatial arrays $I \in \mathbb{R}^{H \times W}$, these methods model boundary tracking through explicit variational and linear estimation formulations, such as the Lucas-Kanade optical flow equation or Kalman filtering.

Because these systems depend on rigid mathematical thresholds rather than statistical learning, their processing throughput is highly efficient, frequently exceeding $150 FPS$ on baseline central processing units (CPUs).

However, this algorithmic rigidity induces critical system failure modes under dynamic clinical conditions:

• Artifact Vulnerability: The presence of specular reflections—which saturate pixel regions to absolute white $Value = 255$—and fast motion blur completely erases the localized spatial edge gradients ∇I required for energy minimization or vector estimation. This causes the tracking loops to collapse entirely.

• Spatial Inflexibility: Out-of-view anomalies require manual, heuristic-driven re-initialization of the tracking seed points.

• Clinical Performance: Consequently, when evaluated on benchmark video datasets such as SUN-SEG, classical methods exhibit poor generalization metrics, yielding a Dice Similarity Coefficient DSC below 0.45 due to high false-positive rates triggered by non-pathological anatomical elements like normal mucosal folds and fecal residue streams.

2. Paradigm 2: Static Deep Learning (Memory-Less)

The advent of deep convolutional neural networks (CNNs) and vision transformers shifted the field toward purely data-driven, spatial-attention architectures. Although deployed on continuous video streams, this paradigm flattens the temporal dimension, mapping a continuous 5D video tensor into an unlinked 4D batch array:

$$X \in \mathbb{R}^{(B \cdot T) \times C \times H \times W}$$

By treating video frames as independent, uncorrelated snapshots, the network completely resets its internal layer activations between successive inference steps, stripping away all temporal context.

While deep 2D encoders and multi-scale attention decoders learn complex structural features—achieving moderate performance profiles $DSC \approx 0.710 - 0.751$on the SUN-SEG dataset—their memory-less nature introduces severe operational instabilities:

• Boundary Discontinuity: When encountering specular highlights or motion blur, the local contrast drops or deforms, causing the network's spatial activations to degrade. This results in continuous mask flickering, internal mask splitting, and complete boundary dropouts.

• Re-entry Latency: Upon an object's re-entry after an out-of-view event, the network exhibits a lagged recovery profile, as it must reconstruct structural confidence from scratch without any historical priors.

• Computational Overhead: Furthermore, processing throughput is highly variable; large transformer-based blocks can drop below $30FPS$ on dedicated graphics hardware, limiting real-time clinical utility.

3. Paradigm 3: Recurrent Spatio-Temporal Networks

Paradigm 3 represents the state-of-the-art in video polyp segmentation by preserving the native 5D sequence tensor dimensionality:

$$X \in \mathbb{R}^{B \times T \times C \times H \times W}$$

This architecture embeds non-linear convolutional recurrent gates—primarily Convolutional Long Short-Term Memory or Gated Recurrent Unit blocks—directly within the highly compressed latent feature bottleneck of a deep neural network.

By passing an internal hidden state vector H_t and cell state vector C_t across successive temporal iterations, the network achieves robust mathematical stability:

• Artifact Suppression: When encountering fast motion blur or specular reflections, the internal forget gates f_t detect the sudden statistical deviation against the historical sequence context. The network suppresses the localized corrupted frames and uses the long-term memory cell to project estimated lesion shapes through the noise block, maintaining smooth, uninterrupted tracking loops.

• Instantaneous Re-localization: Out-of-view anomalies are mitigated by the long-term memory cell, which retains latent structural descriptors of the target adenoma, enabling instant shape re-matching upon re-entry.

• Clinical Superority: By executing these recurrent matrix operations at the highly downsampled bottleneck layer, the framework maintains real-time capabilities $110 FPS$. This spatio-temporal continuity translates to superior performance metrics on the SUN-SEG profile$DSC \approx 0.768 - 0.789$, providing a stable geometric output.

• Failure Modes: System degradation in this paradigm occurs only under compounding long-term tracking errors, where a severely corrupted state matrix propagates geometric distortions forward through the sequence timeline.

REFERENCES

1. Ji, G. P., Chou, Y. C., Fan, D. P., Chen, G., Zhou, T., Van Gool, L., & Shao, L. (2022). Video polyp segmentation: A deep learning perspective. Machine Intelligence Research, 19(6), 531-549. https://doi.org/10.1007/s11633-022-1371-y

2. Almeida, L., & Vasconcelos, M. (2024). PolypNextLSTM: Real-time spatio-temporal video polyp segmentation using lightweight recurrent networks. Computer Methods and Programs in Biomedicine, 246, 108041. https://doi.org/10.1016/j.cmpb.2024.108041

3. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems (NeurIPS), 802-810.

4. Bernal, J., Sánchez, F. J., Vilariño, F., & Javier, F. (2012). Towards automatic polyp detection with a polyp appearance model. Computerized Medical Imaging and Graphics, 36(2), 146-160. https://doi.org/10.1016/j.compmedimag.2011.06.006

5. Fan, D. P., Ji, G. P., Zhou, T., Chen, G., Fu, H., Shen, J., & Shao, L. (2020). PraNet: Parallel reverse attention network for polyp segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 526-536. https://doi.org/10.1007/978-3-030-59713-9_51

6. Liu, X., Mat Isa, N. A., & Lv, F. (2025). Colorectal Polyp Segmentation Based on Deep Learning Methods: A Systematic Review. Journal of Imaging, 11(9), 214. https://doi.org/10.3390/jimaging11090214