**Stainalyzer: A lightweight, open-source toolkit for microscopy image segmentation and staining quantification using Dirichlet masking networks**

*Eduardo Gade Gusmao****1,2****, Neila Caroline Henrique da Silva****1,3****, Débora Rayanne de ArrudaSilva****1****, Nathália Lorena Zeferino Araujo****1****, Norma Lucena Cavalcanti Licinio da Silva****1,#***

**1** Eduardo G Gusmao, Neila CH da Silva, Débora RA Silva, Nathália LZ Araujo and Norma LC Licinio da Silva are with the Oswaldo Cruz Foundation (Fiocruz), Aggeu Magalhaes, Institute, Immunogenetics Department, 50740-465 Recife, Brazil (email: debora.2020208580@unicap.br; email: nathalia.zeferino@ufpe.br; normalucena@fiocruz.com.br). **2** Eduardo G Gusmao is with the Technical University of Munich (TUM), School of Life Sciences, 85354 Freising, Germany (e-mail: eduardo.gusmao@tum-sls.de). **3** Neila CH da Silva is with the Professor Fernando Figueira Institute of Integral Medicine (IMIP), Pediatric Oncology Service, Laboratory of Molecular Biology (e-mail: imip neila.silva@imip.org.br). **#** Corresponding Autor.

**Abstract:**

Automated microscopy image analysis tools are often limited by hardware requirements, narrow modality support, or restrictive commercial licenses. We present Stainalyzer, an open-source toolkit for robust and efficient cellular segmentation and tissue staining quantification across diverse microscopy modalities. Powered by a novel Dirichlet Process-guided masking network architecture, Stainalyzer integrates transfer learning and nonparametric modeling to achieve high accuracy with low computational overhead. We evaluate Stainalyzer on 10 publicly available datasets spanning four microscopy modalities, totaling over 12 million images, and benchmark its performance against 76 competing methods using 29 evaluation and 10 hardware metrics. Stainalyzer supports both CLI and GUI-based workflows, offering a plug-and-play solution for biomedical labs seeking accessible, scalable, and transparent deep learning pipelines for microscopy image analysis.

**Keywords:** Biomedical image processing, Bioinformatics, Deep learning, Transfer learning, Neural networks, Explainable AI, Nonparametric statistics.

**Introduction**

Microscopy is a cornerstone of modern research across disciplines including biology, chemistry, and materials science. To meet diverse experimental needs, a wide range of imaging modalities has emerged – each offering unique strengths. Brightfield microscopy enables visualization of cellular and tissue morphology by capturing transmitted light through stained or naturally contrasted specimens [XXX]. Fluorescence microscopy highlights specific cellular components using targeted dyes or fluorescent proteins that emit light when exposed to particular wavelengths – commonly referred to as imaging “channels” [XXX]. Confocal microscopy improves axial resolution by isolating optical sections, allowing reconstruction of high-contrast three-dimensional structures [XXX]. Lastly, Mass Spectrometry (MS) imaging enables spatial mapping of molecular compositions directly from tissue, without conventional staining, by detecting unique chemical signatures of the target molecules [XXX].

Despite these technological advancements, image analysis remains a largely manual, repetitive, and resource-intensive task in most biological laboratories. Researchers frequently rely on generalized image editing software [XXX] or semi-automated commercial tools [XXX], many of which lack batch-processing capabilities [XXX], offer limited modality support [XXX], and are often closed-source [XXX]. Although deep learning-based pipelines have demonstrated remarkable performance in segmentation and quantification tasks, these tools are commonly hindered by high computational requirements, fragile installation processes, and poorly maintained codebases – making them effectively inaccessible for many laboratories lacking dedicated computational personnel or resources [XXX].

While recent advances in computer vision have centered on increasingly deep and complex architectures [XXX], these models often disregard the practical constraints and workflows of microscopy-based research. In particular, there remains a clear disconnect between what modern software systems offer and what wet-lab researchers require: robust, interpretable, and easy-to-deploy pipelines capable of processing large image batches from diverse microscopy modalities – without reliance on GPUs, cloud platforms, or highly specialized configuration environments [XXX]. Moreover, cost-prohibitive commercial licenses and the opacity of proprietary solutions present additional barriers to widespread adoption.

The increasing complexity of deep learning architectures is, in part, a response to the substantial variability inherent to microscopy imaging. Differences in slide preparation, imaging hardware, staining protocols, and acquisition parameters contribute to significant inter- and intra-modality variability [XXX]. Further challenges arise from morphological variability in biological targets – such as shape, size, intensity, and spatial overlap – which complicate segmentation and quantification tasks [XXX]. For example, in fluorescence microscopy, fluorophores often require different excitation/emission spectra to be captured properly, and overlapping spectra can degrade image quality and limit multiplexing capabilities [XXX]. Together, these factors form a nearly insurmountable barrier to automation, especially when generalization across modalities is desired.

To address these challenges, we introduce Stainalyzer, a computational toolkit for automated microscopy image analysis. At its core lies LumenNet, a lightweight neural engine based on transfer learning and nonparametric modeling, guided by Dirichlet Process-based masking. This architecture enables both: (1) efficient learning of transferable weights across modalities, and (2) drastically reduced training and inference times – achieving performance convergence up to four orders of magnitude faster than current state-of-the-art models [XXX].

Stainalyzer was evaluated on nine public and one internal dataset, encompassing a wide range of use cases across four microscopy modalities and multiple subtypes. These datasets – selected for both diversity and quality – comprise over 12 million image tiles, enabling robust generalizability assessment. To benchmark Stainalyzer, we compared its performance against 67 alternative methods, including 76 alternative methods, including 50 top-performing published models and 26 curated baselines. Evaluation was conducted using 29 segmentation and quantification metrics, along with 10 hardware-specific performance indicators, emphasizing both accuracy and efficiency.

Stainalyzer is the first lightweight, plug-and-play toolkit capable of performing both cell segmentation and staining quantification across diverse microscopy modalities. Powered by a self-supervised engine with color/layer/channel correction logic, it was designed to meet the scalability, transparency, and usability demands of real-world laboratories.

**Related Work**

**Segmentation and Staining in Microscopy**

Automated segmentation of microscopy images is foundational to modern biomedical analysis [XXX]. Early techniques-edge detection, thresholding (e.g., Otsu [XXX]), region growing [XXX], watershed [XXX] – offered simple pipelines but were sensitive to noise and lacked generalizability [XXX]. The introduction of machine learning enabled more flexible rule-based systems and pixel classifiers (e.g., Ilastik [XXX]), though still limited by manual tuning and dataset dependency [XXX]. A major shift occurred with U-Net [XXX], whose encoder-decoder structure [XXX] and skip connections [XXX] became a standard for biomedical segmentation [XXX]. Subsequent innovations such as Cellpose [XXX] and nnU-Net [XXX] emphasized generalist learning and auto-configuration [XXX]. More recent models leverage attention mechanisms and neural architecture search (NAS; [XXX]), yet often demand heavy compute and rely on curated datasets [XXX].

Staining quantification evolved from manual estimation to AI-driven analysis [XXX]. Classical approaches relied on thresholding and color deconvolution [XXX], notably Ruifrok and Johnston’s optical density-based method [XXX]. Toolkits like ImageJ [XXX], QuPath [XXX], and HistomicsTK [XXX] introduced semi-automated ROI quantification [XXX]. Later tools (e.g., DABQuant [XXX], EpidermaQuant [XXX], IHC Profiler [XXX]) standardized workflows but required manual calibration. The deep learning era introduced integrated models like Mermer [XXX], DeepImageJ [XXX], and MyoFuse [XXX], enabling robust stain quantification [XXX]. Emerging techniques even leverage GAN-based virtual staining [XXX], bypassing chemical protocols entirely [XXX]. Despite these advances, generalizability and accessibility remain ongoing concerns [XXX].

**Learning Paradigms for Microscopy**

Transfer learning (TL; [XXX]) and self-supervised learning (SSL; [XXX]) have reshaped microscopy analysis by reducing dependence on large labeled datasets [XXX]. Early TL approaches fine-tuned CNNs (e.g., VGG [XXX], ResNet [XXX]) pretrained on ImageNet [XXX], but domain shift often limited generalization due to microscopy’s distinct texture, color, and scale [XXX]. Nevertheless, TL improved convergence and enabled few-shot applications [XXX]. SSL gained traction as unlabeled microscopy data became abundant [XXX]. Pretext tasks like rotation prediction [XXX], image inpainting [XXX], and contrastive learning (e.g., SimCLR [XXX], MoCo [XXX]) allowed feature learning without manual annotations [XXX]. Domain-specific SSL methods (e.g., uniDINO [XXX], MitoSSL [XXX], color-aware SSL [XXX]) further tailored these strategies to fluorescence imaging and cross-modal transfer [XXX]. Hybrid SSL-TL approaches now address assay variability and low-data regimes with increasing success [XXX].

The reparameterization trick, first introduced in VAEs [XXX], enabled backpropagation through stochastic latent variables by expressing sampling as a differentiable transformation [XXX]. While initially limited to continuous distributions (e.g., Gaussian), the Gumbel-Softmax trick later allowed differentiable sampling from discrete distributions – crucial for latent class modeling in image segmentation [XXX]. Gumbel-based strategies have since appeared in microscopy for uncertain region modeling [XXX], image synthesis [XXX], and structure disentanglement [XXX]. Yet despite their promise, they remain underused [XXX], in part due to training instability [XXX], compute overhead [XXX], and limited support in mainstream deep learning frameworks [XXX].

**Design Needs for Real-World Lab Tools**

While deep learning has revolutionized microscopy image analysis, a critical disconnect remains between methodological innovation and real-world usability [XXX]. Architectures like U-Net [XXX], Cellpose [XXX], and nnU-Net [XXX] have demonstrated exceptional performance but often require GPU acceleration [XXX], extensive configuration [XXX], and programming expertise – barriers for many biology labs [XXX]. Commercial alternatives provide turnkey solutions [XXX] but are prohibitively expensive, with licensing fees reaching tens of thousands of dollars annually [XXX]. Furthermore, generalization remains fragile: pretrained models frequently underperform when exposed to domain shifts in staining [XXX], resolution [XXX], or hardware [XXX].

Staining quantification tools, though more accessible, often rely on static heuristics [XXX], manual calibration [XXX], or opaque statistical models [XXX]. While some recent methods integrate Gumbel-Softmax [XXX] and other reparameterization tricks to model uncertainty [XXX], their practical deployment is hindered by gradient instability, increased computational load [XXX], and limited support in common libraries [XXX].

From a laboratory standpoint, the need is not for maximal abstraction or architectural novelty but for interpretable, fast, and reliable pipelines [XXX]. Most researchers prioritize batch quantification, minimal parameter tuning, and reproducibility over exotic stochastic embeddings or latent-space reconstructions [XXX]. In this context, even the most sophisticated models risk becoming underused if they fail to meet baseline demands for transparency and control [XXX].

To bridge this gap, we introduce Stainalyzer, an open-source, GUI-ready toolkit that prioritizes usability without compromising performance. Powered by LumenNet – a lightweight, modular, self-supervised engine – Stainalyzer emphasizes transparency, reproducibility, and broad modality support, aligning computational sophistication with the real constraints and priorities of biological research environments.

**Methods**

**Overall Pipeline**

Placeholder.

**Image Pre-Processing**

As part of the initial pre-processing pipeline, we first conducted a Scan Dataset for Control operation, where metadata parsing combined with DBSCAN clustering was utilized to isolate control samples and detect outliers, stabilizing downstream color profiles and acquisition noise. Subsequently, a Tiling + Patch Extraction (Smart Tile Creation) step was performed using an Entropy-based method across pre-defined tile sizes and geometric variations, effectively prioritizing tissue-rich regions while minimizing background noise (Supplementary Fig. XXX). To further enhance data quality, an iterative Artifact Removal + Background Subtraction protocol was implemented: leveraging intermediate embeddings from LumenNet outputs, we employed a self-supervised contrastive filtering approach, refined across no more than two passes per figure, guided by an adaptive stop threshold, denoted as the “Cyrilic Buky” (Б) parameter, typically ranging between 1.2 and 1.6. This cyclic improvement strategy progressively optimized artifact exclusion and background uniformity across the dataset.

In the color pre-processing stage, we first applied Color Deconvolution + Unmixing, combining a reconstructed Macenko-Vahadane normalization framework with the novel addition of a random control generator, “Cyrilic Dobro” (Д), followed by post-hoc Ruifrok and Johnston unmixing using pre-normalized matrices for efficient computation (Proof in Supplementary Methods XXX). Subsequently, a comprehensive Color Normalization strategy was introduced by mapping nine color dimensions (HSL, LAB, and xyY) into a compressed RIA (Rotation, Index, Algebraic) surface, achieving lossless chromatic alignment. Non-basic imaging channels underwent optional normalization via Macenko’s, Reinhard’s, and/or Vahadane’s methods, with normalization needs derived systematically from optimal controls established in initial pre-processing (Supplementary Figure XXX). Finally, Contrast & Brightness (Intensity) Normalization was performed in time by averaging global and semi-local histogram equalizations of the Index channel within the RIA space, providing demonstrably superior luminance harmonization across datasets (Supplementary Figure XXX).

In the final pre-processing stage, Channel Normalization + Channel Ordering was performed using percentile-based scaling (1% - 99%) within the RIA space, allowing seamless handling of both standard and non-standard imaging channels, while fixed modality-dependent channel ordering ensured semantic consistency across datasets (Supplementary Fig. XXXX). Subsequently, Geometric Normalization + Alignment + Morphological Correction were applied, utilizing Thin Plate Spline (TPS) transformations for nonlinear warping correction, particularly for advanced fluorescence, confocal, and mass-spectrometry imaging modalities. Data Augmentation (training-only) incorporated lightweight transformations – geometric (rotation, flip, random crop, resize), RIA-based color jittering, and elastic deformation – executed dynamically using Albumentations and Torchvision.transforms libraries. Finally, Filtering + Noise Reduction was achieved via a custom multilateral filtering algorithm, designed to enhance signal-to-noise ratio while preserving tissue morphology, with full procedural details described in Supplemental Methods XXX.

**LumenNet Architecture and Feed-Forward Pass**

Pre-processed microscopy tile initiates its forward traversal through LumenNet via a structured sequence of masking, feature extraction, and contextual refinement. The process begins in the Clustered Mask Layer, where trainable soft masks , updated at each iteration , are applied via Hadamard product to generate masked inputs . This operation preserves spatial geometry while isolating semantically distinct channel-wise patterns. A looped execution scheme (Fig. 1; Supplementary Section XXX) enables these representations to stabilize across iterations without recursion, terminating upon convergence or after a maximum of steps.

Each masked input is independently processed by a lightweight expert network, TinyNN , yielding an embedding . These outputs are vertically stacked to form the intermediate tensor , which is then processed by a Multi-Head Self-Attention (MHSA) mechanism. The MHSA block, composed of parallel attention heads operating in distinct learned subspaces, captures inter-cluster dependencies and relational structure. The result is a globally enriched tensor , encoding context-aware cluster representations for downstream processing.

The contextually augmented representation is flattened and linearly projected into a logit vector through:

where and are learned parameters. These logits serve as unnormalized evidence for each cluster’s relevance and form the interface to the variational DPMM. Once normalized via Gumbel-Softmax, the resulting assignment vector encodes a soft probabilistic mask over clusters, facilitating differentiable routing with inherent interpretability.

Cluster assignments are obtained through the Gumbel-Softmax reparameterization:

The output yields the latent summary vector , which aggregates TinyNN outputs under a smooth, task-adaptive weighting. This reparameterized sampling mechanism ensures end-to-end differentiability across the latent routing pathway (Supplementary Section XXX).

Iteration halts once assignment vectors converge or exceed the loop cap:

Upon termination, the latent vector is directed to a task-specific Head Module – segmentation or staining – completing the forward pass and producing the final output.

**Variational DPMM**

LumenNet embeds a variational Dirichlet Process Mixture Model (DPMM) to perform dynamic, probabilistic clustering of mask filters across loop iterations. This mechanism employs a reparameterized variational posterior to approximate the intractable true posterior , thereby enabling end-to-end gradient-based inference through the use of Gumbel-Softmax relaxation. As such, latent discrete routing decisions are seamlessly integrated into LumenNet’s differentiable looped framework.

Rather than directly marginalizing the latent structure, LumenNet optimizes the Evidence Lower Bound (ELBO):

where the prior is modeled via a truncated stick-breaking process:

in which the concentration parameter controls cluster sparsity – low encourages specialization, while high promotes balanced usage.

The final component is completed via:

ensuring proper normalization. These weights act as unsupervised routing coefficients, directing feature flow toward semantically coherent submodules and enabling interpretable, data-driven specialization (see Supplementary Section XXX).

To facilitate tractable optimization, a mean-field approximation is employed over latent variables:

with and . The full ELBO is expressed as:

This decomposition enables efficient gradient-based optimization, as many components admit closed-form solutions under the exponential family.

For non-analytic expectations, LumenNet employs the Gumbel-Softmax trick to approximate categorical sampling while maintaining differentiability:

where reflects relative cluster importance, introduces stochasticity, and modulates the entropy of the resulting distribution. Annealing during training ensures a transition from exploratory soft routing to sharp, discrete-like decisions.

To prevent premature cluster collapse, LumenNet includes a KL divergence term, approximated by:

where . During early training, when assignments are diffuse, this expression simplifies to an entropy-based surrogate , promoting diversity and regularization in routing behavior (see Supplementary Section XXX).

Altogether, the attention-refined logits , soft assignments , and truncated stick-breaking prior form a fully reparameterized variational DPMM. This module – executed per iteration – refines cluster responsibilities in a differentiable manner, thereby closing the loop between stochastic inference and deterministic feature flow. The result is a lightweight yet expressive mechanism for unsupervised filter allocation, aligned with both task supervision and architectural interpretability.

**Loss Functions and Backpropagation**

LumenNet’s training pipeline is anchored by a task-specific supervision loss , computed at the output of the Head Layer. Backpropagation initiates exclusively from the final loop iteration , thereby reducing redundancy while preserving semantic alignment. To structure this process, we delineate four distinct gradient streams: (i) task-driven supervision, (ii) variational gradients from the ELBO, (iii) reparameterization through Gumbel-Softmax, and (iv) attention-based projection feedback. These parallel yet complementary streams collectively enable interpretable and modular backpropagation across LumenNet’s hybrid stochastic-deterministic architecture.

In Stream 1, gradients from propagate through the Aggregation Layer, where:

This stream ensures that each TinyNN is updated in proportion to its contribution to the final prediction, enforcing localized specialization and predictive consistency.

The total objective function augments the task loss with a regularization term derived from variational inference:

where governs the influence of latent structural priors during optimization.

In Stream 2, gradients of the KL term are propagated to modulate soft cluster assignments , enforcing probabilistic coherence:

This regularization term encourages diversity among cluster usages while preserving task relevance.

Stream 3 transmits gradients through the reparameterized categorical sampling path:

facilitating smooth gradient propagation despite the presence of sampling-based latent decisions.

Stream 4 begins at the projection logits , backpropagating through the Preparation Layer and MHSA to the original cluster outputs:

This stream completes the circular variational feedback loop, ensuring attention-informed updates to the TinyNNs (see Supplementary Section XXX).

The net update to each TinyNN output aggregates both the task and variational contributions:

merging predictive supervision and latent structure optimization into a single coherent gradient.

Thanks to stream-level separability, LumenNet permits efficient parallel gradient computation:

allowing asynchronous updates across task and variational components. Extensive benchmarking identified AdamW as the preferred optimizer for its convergence stability in high-entropy regimes (Supplementary Fig. XXX).

Through four complementary streams of backpropagation, LumenNet unifies task supervision, latent regularization, attention-based feature flow, and reparameterized inference under a lightweight, modular, and interpretable architecture. This configuration – distinct from conventional Mixture-of-Experts or gated transformer models – enables dynamic clustering and specialization with reduced overhead and robust generalization.

**Image Post-Processing**

In the initial post-processing stage, Mask Filtering + Artifact Removal was performed using lightweight operations – such as erosion, dilation, opening/closing, boundary filtering, and shape filtering – to refine mask borders and eliminate biologically implausible structures; during this phase, critical Б and Д parameters were also computed. Instance Separation was achieved through a fast Watershed transformation applied as a correction factor across tile and image-level outputs, enabling discrete object delineation in densely packed tissue regions. Finally, Contour Smoothing & Edge Correction was conducted by applying Fourier-based contour filtering, selectively preserving low-frequency boundary components to enhance morphological realism and ensure anatomically plausible segmentations.

In the final post-processing stage, Label Normalization & Conflict Resolution leveraged the explainability of LumenNet by revisiting applied filters, resolving over XX% of label conflicts efficiently (Supplementary Fig. XXX). Quantitative Measurement & Extraction Statistics were then performed to compute morphometric, intensity-based, and spatial features, forming the comprehensive dataset utilized by Stainalyzer, with updates to the previously calculated Б and Д parameters. Finally, Coordinate & Metadata Export was executed through flexible, multi-format output options – including CSV, JSON, and GeoJSON – enabling structured dissemination of spatial, biological, and model-specific metadata for seamless downstream integration.

**Experiments**

Placeholder, placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder.

**Results**

Placeholder, placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder.

**Discussion**

Placeholder, placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder.

**Conclusion and Future Investigations**

Placeholder, placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder placeholder.