Representation Learning from Superpixels: Lightweight CNN Encoders Under Data Scarcity

Felipe C. R. Salvagnini[1(✉)][0000−0002−7896−0058], Maria A. K. Miranda[1][0009−0003−7811−9116], Gilson J. Soares[1][0000−0001−8805−5644], Mateus O. Silva[1][0009−0001−5404−7168], Cid Santos[2][0000−0002−9278−5356], Jeova F. S. R. Neto[3][0000−0001−8103−7937], Silvio J. F. Guimarães[4][0000−0001−8522−2056], Oge Marques[5][0000−0003−4321−2719], and Alexandre X. Falcão[1][0000−0002−2914−5380]

[1] Institute of Computing, Universidade de Campinas
{felipe.salvagnini,gilson.soares,maria.miranda,mateus.silva}@students.ic.unicamp.br, afalcao@unicamp.br
[2] Eldorado Research Institute cid.santos@eldorado.org.br
[3] Department of Computer Science, Bowdoin College j.farias@bowdoin.edu
[4] Pontifical Catholic University of Minas Gerais sjamil@pucminas.br
[5] NVIDIA omarques@nvidia.com

Abstract. Effective representation learning in medical imaging is challenging when both labeled data and computational resources are limited. Feature Learning from Image Markers (FLIM) lets users design efficient convolutional encoders by drawing markers on discriminative regions of a few images; the markers act as weak supervision, and filter coefficients are learned directly from feature patches centered at marker pixels without backpropagation. However, reliance on the user affects reproducibility and scalability. Replacing markers with superpixel centers removes the user from the loop but inflates filter count proportionally to training images and superpixels. We propose Superpixel Filter Learning (SPiFiL), which learns filters from superpixel centers and retains only the most discriminative ones via intra- and inter-class rankings. To showcase robustness, we select three complementary parasitology classification datasets, a critical problem in developing countries that requires an effective solution under low-resource settings. By learning encoder filters from one image per class, SPiFiL can reduce the filter count by up to 25× while achieving competitive performance and improvements of up to +0.032 in Cohen's kappa and +9.2% in F1-score (using a linear SVM classifier). Compared with strong baselines, SPiFiL achieves competitive performance using only a 0.0525M-parameter encoder against models ranging from 1.2M (SqueezeNet) to 134.3M (VGG16), whether trained from scratch or pretrained on ImageNet, highlighting its potential for scalable, low-resource medical imaging.

Keywords: Feature Learning from Image Markers · Superpixels · Filter learning and selection · Data scarcity · Resource-constrained learning.

1 Introduction

Over the past years, the medical image analysis community has witnessed the broad acceptance of deep learning methods [10] across problems such as image classification, segmentation, and synthesis. Nevertheless, this acceptance is overstated by reports on well-curated benchmarks [10, 11] — settings in which pre-trained models can be easily fine-tuned and where abundantly annotated data are available. However, scarce data annotation and limited computational resources are critical constraints in practice, especially in developing countries. Under these constraints, large pre-trained models are too costly to deploy, and compact models trained from scratch fail to learn. A representative example is the detection of intestinal parasites in microscopy images [1, 21], where both constraints compound simultaneously. Then, *how to learn efficient convolutional encoders with limited annotated data*, such that the resulting feature space allows a classifier to separate each class from the others reliably? From the perspective of representation learning [4], meaningful feature representations are key to successful downstream tasks, yet most approaches still depend on large datasets and costly pretraining schemes, such as self-supervision [7, 20], leaving the extremely low-data regime largely underexplored [16].

FLIM (Feature Learning from Image Markers) [12, 17] leverages human knowledge to extract image patches from discriminative regions of a few weakly annotated training images as priors to estimate filter coefficients for convolutional encoders. Specifically, filter coefficients are estimated from image patches centered on marker pixels: patches from the first layer define the filters directly, and their spatial locations are propagated to subsequent layers to extract patches in the transformed feature space. The key insight is that a convolution measures the similarity between a given visual pattern (*i.e.,* filter coefficients) and all image patches, thereby enhancing discriminative regions. Unlike few-shot learning [13, 18], which relies on large meta-training corpora and episodic training, or scribble-based methods [19, 22], FLIM estimates filter coefficients directly from a few user-marked images [8, 15], without backpropagation or external data.

Despite its effectiveness, existing FLIM-based methods suffer from two key limitations: *subjectivity* and *scalability*. Since the user chooses both the images to annotate and the discriminative regions to manually draw markers (seed pixels), performance remains sensitive to the user's expertise. Moreover, each image must be marked individually, making annotation impractical as the number of classes or the required number of images grows (*e.g.,* 10 or more classes). This work addresses both limitations by automating filter learning into a method we term *Superpixel Filter Learning* (SPiFiL)[1]: images are randomly selected from the training set, and user-drawn seeds are replaced with superpixel geodesic centers. Using a single image per class, we achieve results comparable to user-based approaches; however, this strategy significantly increases the filter count (*e.g.,* 50 superpixels on an 8-class problem yields 400 filters per layer). To tackle this, SPiFiL introduces a patch-ranking strategy based on intra- and inter-class simi-

[1] https://github.com/LIDS-UNICAMP/SPiFiL.

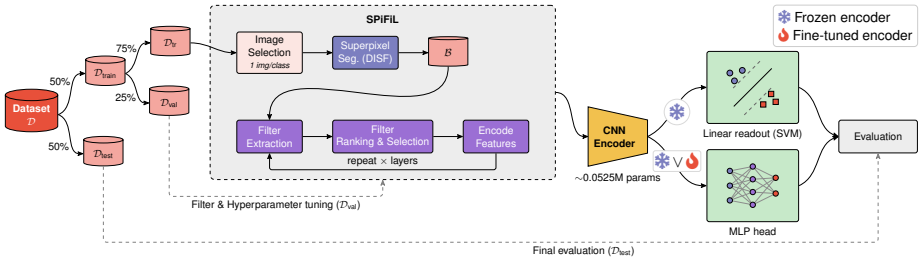


Fig. 1. Overview of the SPiFiL pipeline: per-layer superpixel filter extraction, ranking, selection, and feature extraction, feeding a downstream SVM or MLP head.

larities to retain only the most discriminative filters — reducing the filter count by up to 25× (*e.g.,* 400 → 16) while improving classification on 2 of 3 datasets (up to +9.2% in F1-score).

2 Method & Theoretical Foundations

Figure 1 depicts SPiFiL. From training images with dataset-provided semantic masks (no extra annotation), randomly sampled per class, a superpixel algorithm partitions each image and the geodesic center of every superpixel defines a candidate filter; all candidates across images form the layer's filter bank. Intra- and inter-class similarities then rank and select the most discriminative filters, which encode features (pooling with stride s) for the next layer. Repeating this per layer yields the encoder, after which a downstream classifier (*e.g.,* SVM, MLP) is trained on the final features.

Algorithm 1 formalizes the method. We briefly revisit FLIM, then detail our contribution — superpixel-based automated filter learning with discriminative selection — referring to previous works for a fuller FLIM exposition [8, 12, 14, 15, 17].

2.1 Feature Learning From Image Markers (FLIM)

Given a 2D image $\hat{I} = (D_I, \mathbf{I})$, where each pixel $p \in D_I \subset \mathbb{Z}^2$ holds a feature vector $\mathbf{I}(p) \in \mathbb{R}^m$, a patch $\mathbf{P}(p) \in \mathbb{R}^{k \times k \times m}$ is defined as the concatenation of feature vectors in a $k \times k$ neighborhood centered at p. This work builds on the Bag of Feature Points [12], a variant of FLIM, which identifies discriminative spatial locations $\mathcal{B}$ via single-pass K-means clustering on patches extracted from user-marked pixels, then reuses those locations by projection across all encoder blocks. At each block, patches at locations in $\mathcal{B}$ are z-score normalized and used to define kernels $\mathbf{W}_i = \mathbf{P}(p)/|\mathbf{P}(p)| \oslash \boldsymbol{\sigma}_B$ and biases $\mathbf{b}_i = -\langle \boldsymbol{\mu}_B, \mathbf{W}_i \rangle$, where $\boldsymbol{\mu}_B$ and $\boldsymbol{\sigma}_B$ are the mean and standard deviation of the patch set $\mathcal{P}_B$, and $\oslash$ denotes element-wise division.

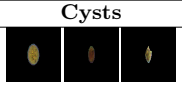
2.2 Superpixel-based feature learning

SPiFiL eliminates subjectivity by replacing user-drawn markers with an automatic superpixel segmentation step, dispensing with the K-means clustering of BoFP [12] entirely. We adopt DISF [3], a seed-based method that starts from an oversampled grid of $n_0 \gg n_s$ seeds and iteratively reduces them to the desired n_s superpixels by discarding the least relevant ones. In a comprehensive benchmark of superpixel methods [2], DISF ranks among the best according to several metrics while offering precise control over the number of superpixels, fixing a deterministic per-image candidate budget of n_s seeds [3]. DISF runs within the ground-truth semantic mask, so every superpixel center — and thus every candidate filter — lies on the class of interest, not on background or impurities. The geodesic center of each superpixel τ_x — defined as its medoid $p_x^* = \arg\min_{p \in \tau_x} \|\mathbf{I}(p) - \boldsymbol{\mu}_{\tau_x}\|_2$ — serves as a direct substitute for user-drawn markers. The union of all such centers across randomly selected images defines the patch candidate set $\mathcal{B} = \bigcup_{\hat{I} \in \mathcal{D}_T} \{p_x^* \mid \tau_x \in \mathrm{DISF}(\hat{I}, M)\}$, motivating the filter selection described next.

2.3 Filter Selection

Replacing user-drawn markers with superpixel centers yields a filter bank whose size scales with the total number of superpixels across all training images — far exceeding the n_f filters desired per layer. We address this with a two-stage selection strategy: (i) a per-class discriminability score ranks every candidate filter, and (ii) a diversity-aware greedy selection retains n_f filters that are both discriminative and mutually diverse.

Table 1. Dataset Composition (genus abbreviated; percentages are class shares)

Eggs		**Larvae**		**Cysts**	
					
H. nana	348 (6.8%)	*S. stercoralis*	446 (12.7%)	*E. coli*	719 (7.5%)
H. diminuta	80 (1.6%)	Impurities	3068 (87.3%)	*E. histolytica*	78 (0.8%)
Ancylostoma	148 (2.9%)	**Total**	**3514**	*Endolimax nana*	724 (7.6%)
E. vermicularis	122 (2.4%)			*G. intestinalis*	641 (6.7%)
A. lumbricoides	337 (6.6%)			*I. bütschlii*	1501 (15.7%)
T. trichiura	375 (7.3%)			*B. hominis*	189 (2.0%)
S. mansoni	122 (2.4%)			Impurities	5716 (59.7%)
Taenia spp	236 (4.6%)			**Total**	**9568**
Impurities	3344 (65.4%)				
Total	**5112**				

Per-class scoring and selection. Let $\mathcal{F}$ denote the set of candidate patches extracted at superpixel centers across training images, and $\boldsymbol{\Sigma}^{-1}$ a shared Mahalanobis metric, taken as the (inverse) empirical covariance estimated over $\mathcal{F}$ [9].[2]

[2] When $\boldsymbol{\Sigma}$ is singular, we fall back to Euclidean distance.

Algorithm 1 Automated FLIM via Superpixel Seeds and Filter Selection

Input: Images $\mathcal{I}$ with masks $\mathcal{M}$ and labels $\mathcal{C}$; layers N; superpixels n_s; filters n_f; kernel size K; stride s; diversity weight α; pool factor γ.
Output: Kernels $\{\mathbf{W}^{(l)}\}_{l=1}^{N}$, biases $\{\mathbf{b}^{(l)}\}_{l=1}^{N}$.
 # Layer 0: seed extraction
for $\forall\, I_i \in \mathcal{I}$ **with** mask $M_i \in \mathcal{M}$, label $\ell_i \in \mathcal{C}$ **do**
  $S_i^{(0)} \leftarrow \text{SuperpixelCenters}(\text{DISF}(\text{LABNorm}(I_i),\, M_i,\, n_s),\, \ell_i)$
end for
Score all seeds by the Fisher ratio; assign per-class ranks h
 # Layer loop
for $l \leftarrow 1$ **to** N **do**
  $n_c \leftarrow \lfloor n_f / |\mathcal{C}| \rfloor$ ▷ Filters per class
  $S^* \leftarrow \bigcup_c \text{DiverseSelect}_c(\gamma n_c,\, n_c,\, \alpha)$ ▷ $\alpha\hat{\phi} + (1-\alpha)\delta$
  Build kernels $\mathbf{W}^{(l)}, \mathbf{b}^{(l)}$ from patches at S^* ▷ See Sec. 2.1
  for $\forall I_i \in \mathcal{I}$ **do**
    $\hat{I}_i^{(l)} \leftarrow \text{MaxPool}(\text{ReLU}(\mathbf{W}^{(l)} \hat{I}_i^{(l-1)} + \mathbf{b}^{(l)}),\, s)$
    $S_i^{(l)} \leftarrow \text{ProjectSeeds}(S_i^{(l-1)},\, s)$
  end for
  if $l < N$ **then** re-score $S_i^{(l)}$ and update ranks
  end if
end for
return $\{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}_{l=1}^{N}$

For each class $c \in \mathcal{C}$ independently, every candidate s_j with label $\ell_j = c$ receives a Fisher-ratio score [6]: $\phi(j) = \bar{d}_{\text{inter}}(j)/(\bar{d}_{\text{intra}}(j) + \varepsilon)$, where $\bar{d}_{\text{intra}}$ and $\bar{d}_{\text{inter}}$ are the mean Mahalanobis distances from s_j to same- and different-class samples; ranking per class (rather than globally) yields a rank h_j within c and prevents classes with intrinsically higher scores from monopolizing the top positions. To avoid selecting near-duplicate filters, we then form a pool of the $\gamma\, n_c$ top-ranked seeds per class ($n_c = \lfloor n_f/|\mathcal{C}| \rfloor$, $\gamma \geq 1$) and greedily select n_c of them, each chosen to maximize $\alpha\, \hat{\phi}(j) + (1-\alpha)\, \delta(j)$ — trading off the normalized Fisher score $\hat{\phi}(j) = 1 - (h_j - 1)/\max(h_c)$ against the Euclidean distance $\delta(j)$ to the already-selected set, with $\alpha \in [0, 1]$ setting the discriminability–diversity balance. Filters for layer l are built from the selected patches (Section 2.1); the layer then encodes all images and the seed coordinates are projected to the new feature domain, where scores are recomputed before selecting filters for layer $l + 1$, since superpixels are computed only once but seed discriminability shifts after each encoding step.

3 Experiments & Results

Datasets. We evaluate on three publicly available datasets collected according to parasitology protocols, each split into 50%/50% training/test ($\mathcal{D}_{\text{train}}/\mathcal{D}_{\text{test}}$) across 3 stratified splits (repeated holdout, not 3-fold CV). From each training split, a single image is randomly selected to learn the encoder for SPiFiL and all

Table 2. Encoder + SVM performance across variants and datasets ($\alpha=0.5$, $\gamma=5$)

Dataset	Technique (Filter Count)	Acc. (%)	F1-Score (%)	Kappa
Eggs	All Pixels [200, 200, 200]	94.66 ± 0.29	91.53 ± 0.50	0.9049 ± 0.0051
	Center Pixel [40, 40, 40]	93.85 ± 0.99	90.05 ± 1.87	0.8899 ± 0.0178
	50 Superpixels [400, 400, 400]	$\mathbf{94.76 \pm 0.70}$	91.40 ± 0.81	$\mathbf{0.9062 \pm 0.0128}$
	SPiFiL [16, 32, 48]	93.70 ± 0.95	$\mathbf{93.76 \pm 0.91}$	0.8887 ± 0.0159
Larvae	All Pixels [100, 100, 100]	96.76 ± 0.49	92.65 ± 1.20	0.8531 ± 0.0240
	Center Pixel [20, 20, 20]	96.26 ± 0.43	91.66 ± 0.91	0.8333 ± 0.0182
	50 Superpixels [100, 100, 100]	96.55 ± 0.19	92.28 ± 0.46	0.8455 ± 0.0092
	SPiFiL [16, 32, 48]	$\mathbf{97.06 \pm 0.14}$	$\mathbf{97.05 \pm 0.14}$	$\mathbf{0.8666 \pm 0.0059}$
Cysts	All Pixels [150, 150, 150]	87.60 ± 0.78	80.71 ± 1.49	0.7973 ± 0.0123
	Center Pixel [30, 30, 30]	86.95 ± 0.16	79.45 ± 0.44	0.7869 ± 0.0021
	50 Superpixels [300, 300, 300]	87.86 ± 0.30	79.75 ± 0.66	0.8015 ± 0.0051
	SPiFiL [12, 30, 48]	$\mathbf{89.83 \pm 0.74}$	$\mathbf{89.91 \pm 0.71}$	$\mathbf{0.8332 \pm 0.0115}$

FLIM-based methodologies; SPiFiL's hyperparameters — the number of filters n_f, the diversity weight α, and the pool factor γ — are selected on a 75%/25% training/validation ($\mathcal{D}_{\text{tr}}/\mathcal{D}_{\text{val}}$) partition of $\mathcal{D}_{\text{train}}$. To evaluate classification under different data regimes, we further sample 1%, 5%, 25%, 50%, 75%, and 100% of the training subset, using the remainder for validation (for 100%, training and validation sets coincide). Table 1 summarizes the dataset compositions.

Encoder architecture. All FLIM-based and comparison methods share a three-layer convolutional encoder: each layer applies 5×5 convolution, followed by ReLU activation and 3×3 max-pooling with stride 2.

FLIM baselines (no backpropagation). We first evaluate the original FLIM pipeline. Two user-based marker strategies are considered: (1) *All Pixels*, in which the markers are all pixels within a drawn click region, and (2) *Center Pixel*, which uses only the center pixel of each click. The extracted features train/test a linear SVM (one-vs-one, $C=100$).

Automated filter learning. We replace user-drawn markers with DISF superpixel geodesic centers [3] ($n_s = 50$ superpixels per image), using the same randomly selected images. It yields $50 \times |\mathcal{C}|$ candidate filters per layer.

Filter selection and ranking ablation. Following Section 2.3, we first fix the number of filters per layer on the validation set ($n_f = (16, 32, 48)$, from $n_f \in \{8, 16, 32, 48\}$), then ablate the ranking hyperparameters at this n_f: diversity weight $\alpha \in \{0.3, 0.5, 0.7, 1.0\}$ and pool factor $\gamma \in \{1, 3, 5\}$, scored via SVM on validation only. Figure 2 shows the ablation results on validation κ. Performance is stable, indicating robustness to hyperparameters. Diversity drives the gains: optima lie at $\alpha<1.0$, so pure Fisher ranking ($\alpha=1.0$) is never optimal. The pool factor matters too: a larger candidate pool ($\gamma \geq 3$) consistently outperforms $\gamma=1$. We adopt $\alpha=0.5$, $\gamma=5$ throughout. Table 2 reports the adopted configuration trained on the full training set and tested on the held-out set.

Neural network comparison. To assess SPiFiL-learned encoders, we attach an MLP classification head: `AdaptiveAvgPool2d(1)`, followed by two hidden layers (256 and 128 units, ReLU, dropout 0.3) and a linear output layer. We

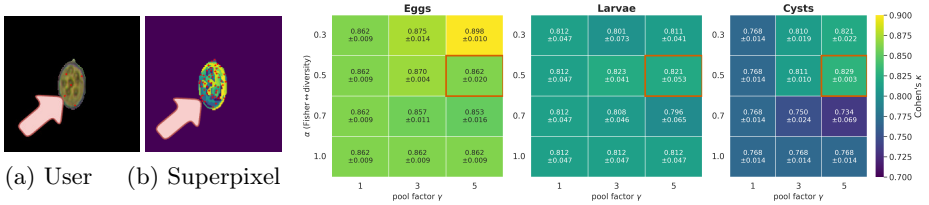


Fig. 2. (a,b) User markers vs. uniform superpixel seeds (red arrows); (right) ranking ablation: validation κ over α and γ, with the adopted configuration boxed.

compare the following strategies on the same encoder architecture: (1) *SPiFiL (Frozen)*: SPiFiL-learned encoder weights are fixed; only the MLP head is trained; (2) *SPiFiL (Fine-tuning)*: the full network (encoder + MLP) is fine-tuned end-to-end; (3) *He*: encoder weights are randomly initialized (He) and trained end-to-end. All NN experiments use Adam (lr 10^{-3}, weight decay 10^{-4}, batch 32), `ReduceLROnPlateau` (factor 0.5, patience 10), and early stopping (patience 50 on validation loss, max 300 epochs); we report accuracy, weighted F1, and Cohen's κ [5].

Automated vs. user-based selection. Table 2 compares two user-based FLIM strategies (*All Pixels*, *Center Pixel*) against two automated variants: superpixel-based marker placement (*50 Superpixels*) and *SPiFiL* full automation. Only superpixel-based selection is competitive with both user-based strategies on all three datasets, confirming that automating marker placement does not sacrifice discriminability. SPiFiL further improves Larvae and Cysts across all metrics while using far fewer filters: user-based methods produce as many filters per layer as markers (growing with classes and clicks), whereas SPiFiL fixes the budget to $n_f = (16, 32, 48)$ regardless of problem complexity (Figure 2). For the user-based baselines, we use 5 clicks per parasite class for Eggs and Cysts (8 and 6 classes, excluding impurities) and 10 clicks per class for Larvae (2 classes). This asymmetry is intentional: experts place few markers in discriminative regions, whereas superpixels provide many candidates for post-hoc selection.

Eggs class imbalance. On Eggs, SPiFiL shows a slight drop in accuracy/κ and an increase in weighted F1 ($91.4 \rightarrow 93.6$). With 8 parasite classes, Eggs has the most compressed bank (400 filters/layer). Since filter learning excludes impurity patches, the retained filters favor inter-parasite discrimination, improving per-class balance at a small cost to the dominant Impurities class (65.4%).

SPiFiL Encoder Benchmark. SPiFiL-learned encoders consistently outperform randomly initialized (He) counterparts across all datasets and training fractions (Figure 3), most markedly at 1–5% training data, where He fails to learn meaningful representations, while SPiFiL already achieves competitive kappa scores. With only a $\leq$52.5K-parameter encoder ($\leq$99K with the MLP head) — orders of magnitude fewer than standard CNNs — the SPiFiL encoder followed by a linear SVM or MLP head achieves competitive performance while remaining deployable on resource-constrained hardware. Learning the encoder is a one-time, backpropagation-free *training* step: on a single CPU thread (Intel Core i9-

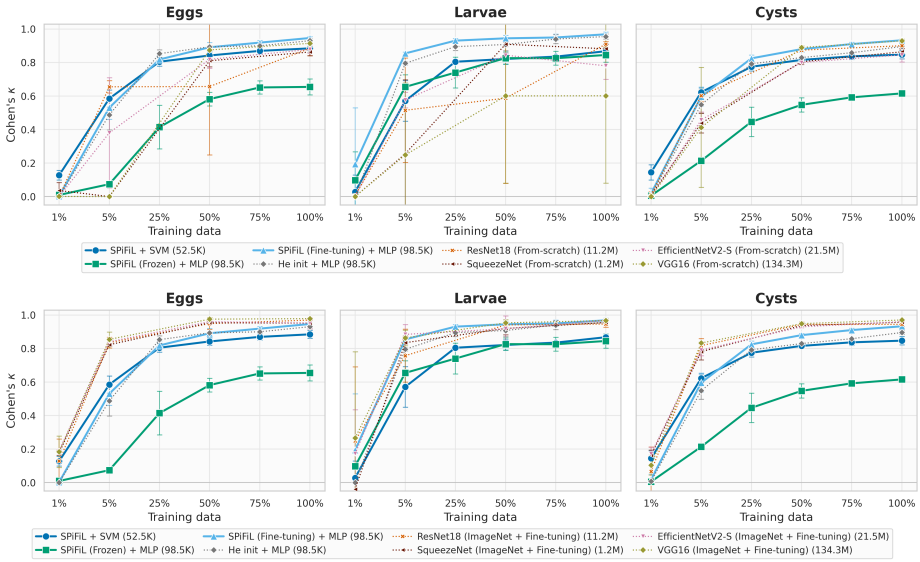


Fig. 3. SPiFiL vs. standard CNN baselines: Cohen's κ across training fractions.

14900KF, no GPU) — a deliberately conservative low-resource setting — DISF costs ∼20 ms/image, and filter learning and selection take 1.3/12.9/27.1 s for Larvae/Cysts/Eggs, dominated by the Fisher scoring (scales with the class count) and reduced ∼2.5× by OpenMP threads (*e.g.,* Eggs 27→11 s). Once learned, *inference* is negligible, as the encoder needs only ≤0.54 GFLOPs (∼28 ms/image). When the encoder is unfrozen, end-to-end fine-tuning further improves performance, achieving the highest kappa on Eggs and Cysts at most training fractions. Remarkably, SPiFiL+SVM sometimes matches backpropagation variants even on 100% of the data, indicating well-separated features.

Deep learning baselines. We further compare SPiFiL against four standard CNN architectures — ResNet18, SqueezeNet, EfficientNetV2-S, and VGG16 (1.2M–134.3M parameters) — each trained both from scratch and fine-tuned from ImageNet weights on the same splits and fractions (Figure 3). Trained from scratch, all four collapse at 1% ($\kappa \approx 0$) and remain unstable at 5%, whereas SPiFiL, initialized from a single image per class without backpropagation, already yields usable encoders. ImageNet-pretrained backbones are strongest overall (*e.g.,* VGG16 $\kappa \approx 0.97$ at 100%) but rely on large-scale pretraining and millions of parameters. In contrast, SPiFiL stays competitive at a fraction of the cost, narrowing the gap where annotation and compute are scarcest.

4 Conclusion

This work introduced SPiFiL, an automated, superpixel-driven method for learning CNN filters that eliminates the subjectivity inherent in user-drawn markers.

Its ranking strategy cuts the filter count by up to 25× while improving classification on 2 of 3 datasets, and the encoder — from a *single image per class* — outperforms randomly initialized networks. That a linear SVM on frozen SPiFiL features can match the performance of end-to-end fine-tuning shows that it learns a discriminative, linearly separable representation without backpropagation. At ≤52.5K encoder parameters (≤99K with the head), it is well-suited to parasite screening where data, compute, and infrastructure are scarce.

The single-image-per-class setting isolated the contributions of automated selection and ranking; future work will lift it (larger subsets, filter-level augmentation, impurity-class patches), extend SPiFiL to other medical domains and 3D data, and ultimately remove the dependence on ground-truth masks.

Acknowledgments. We thank the Eldorado Research Institute. This work was financially supported by CNPq (304711/2023-3, 309593/2026-3, 408003/2025-1, and 442950/2023-3), FAPEMIG (APQ-01079-23 and APQ-03964-25), FAPESP (2023/14427-8, 2025/23020-4, and 2025/19856-0) and CAPES (STIC-AMSUD 88887.878869/2023-00).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anantrasirichai, N., Chalidabhongse, T.H., Palasuwan, D., Naruenatthanaset, K., Kobchaisawat, T., Nunthanasup, N., Boonpeng, K., Ma, X., Achim, A.: ICIP 2022 challenge on parasitic egg detection and classification in microscopic images: Dataset, methods and results. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 4306–4310 (2022). https://doi.org/10.1109/ICIP46576.2022.9897267
2. Barcelos, I.B., Belém, F.D.C., João, L.D.M., Patrocínio, Z.K.G.D., Falcão, A.X., Guimarães, S.J.F.: A comprehensive review and new taxonomy on superpixel segmentation. ACM Comput. Surv. **56**(8) (Apr 2024). https://doi.org/10.1145/3652509, https://doi.org/10.1145/3652509
3. Belém, F.C., Guimarães, S.J.F., Falcão, A.X.: Superpixel segmentation using dynamic and iterative spanning forest. IEEE Signal Processing Letters **27**, 1440–1444 (2020). https://doi.org/10.1109/LSP.2020.3015433
4. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(8), 1798–1828 (2013). https://doi.org/10.1109/TPAMI.2013.50, http://ieeexplore.ieee.org/document/6472238/
5. Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement **20**(1), 37–46 (1960)
6. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**(2), 179–188 (1936)
7. Huang, S.C., Pareek, A., Jensen, M., Lungren, M.P., Yeung, S., Chaudhari, A.S.: Self-supervised learning for medical image classification: a systematic review and implementation guidelines. npj Digital Medicine **6**(1), 74 (2023). https://doi.org/10.1038/s41746-023-00811-0, https://www.nature.com/articles/s41746-023-00811-0

8. Joao, L.d.M., Santos, B.M.d., Guimaraes, S.J.F., Gomes, J.F., Kijak, E., Falcao, A.X., et al.: A flyweight CNN with adaptive decoder for Schistosoma mansoni egg detection. arXiv preprint arXiv:2306.14840 (2023). https://doi.org/10.48550/arXiv.2306.14840
9. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis **88**(2), 365–411 (2004). https://doi.org/10.1016/S0047-259X(03)00096-4, https://www.sciencedirect.com/science/article/pii/S0047259X03000964
10. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005, https://www.sciencedirect.com/science/article/pii/S1361841517301135
11. Maier-Hein, L., Eisenmann, M., Reinke, A., Onogur, S., Stankovic, M., Scholz, P., Arbel, T., Bogunovic, H., Bradley, A.P., Carass, A., Feldmann, C., Frangi, A.F., Full, P.M., Van Ginneken, B., Hanbury, A., Honauer, K., Kozubek, M., Landman, B.A., März, K., Maier, O., Maier-Hein, K., Menze, B.H., Müller, H., Neher, P.F., Niessen, W., Rajpoot, N., Sharp, G.C., Sirinukunwattana, K., Speidel, S., Stock, C., Stoyanov, D., Taha, A.A., Van Der Sommen, F., Wang, C.W., Weber, M.A., Zheng, G., Jannin, P., Kopp-Schneider, A.: Why rankings of biomedical image analysis competitions should be interpreted with care. Nature Communications **9**(1), 5217 (2018). https://doi.org/10.1038/s41467-018-07619-7, https://www.nature.com/articles/s41467-018-07619-7
12. Martinelli, J.D., Rodrigues Filho, M.L., Salvagnini, F.C.d.R., Soares, G.J., Santos, J.A.d., Falcão, A.X.: *flim* networks with bag of feature points. In: Chaves, D., Forero Vargas, M., Rojas Camacho, O. (eds.) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. pp. 268–281. Springer Nature Switzerland, Cham (2026)
13. Pachetti, E., Colantonio, S.: A systematic review of few-shot learning in medical imaging. Artificial Intelligence in Medicine **156**, 102949 (2024). https://doi.org/10.1016/j.artmed.2024.102949, https://www.sciencedirect.com/science/article/pii/S093336572400191X
14. Salvagnini, F.C.d.R., Gomes, J.F., Santos, C.A.N., Guimarães, S.J.F., Falcão, A.X.: Improving FLIM-based salient object detection networks with cellular automata. In: 2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 1–6 (2024). https://doi.org/10.1109/SIBGRAPI62404.2024.10716266
15. Soares, G.J., Cerqueira, M.A., Guimaraes, S.J.F., Gomes, J.F., Falcão, A.X.: Adaptive decoders for FLIM-based salient object detection networks. In: 2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 1–6 (2024). https://doi.org/10.1109/SIBGRAPI62404.2024.10716333
16. Song, Y., Wang, T., Cai, P., Mondal, S.K., Sahoo, J.P.: A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Comput. Surv. **55**(13s) (Jul 2023). https://doi.org/10.1145/3582688, https://doi.org/10.1145/3582688
17. de Souza, I.E., Falcão, A.X.: Learning CNN filters from user-drawn image markers for coconut-tree image classification. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022). https://doi.org/10.1109/LGRS.2020.3020098
18. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: A survey on few-shot learning. ACM Comput. Surv. **53**(3) (Jun 2020). https://doi.org/10.1145/3386252, https://doi.org/10.1145/3386252

19. Wei, Y., Ji, S.: Scribble-Based Weakly Supervised Deep Learning for Road Surface Extraction From Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–12 (2022). https://doi.org/10.1109/TGRS.2021.3061213, https://ieeexplore.ieee.org/document/9372390/
20. Wolf, D., Payer, T., Lisson, C.S., Lisson, C.G., Beer, M., Götz, M., Ropinski, T.: Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging. Scientific Reports **13**(1), 20260 (2023). https://doi.org/10.1038/s41598-023-46433-0, https://www.nature.com/articles/s41598-023-46433-0
21. Zhang, C., Jiang, H., Jiang, H., Xi, H., Chen, B., Liu, Y., Juhas, M., Li, J., Zhang, Y.: Deep learning for microscopic examination of protozoan parasites. Computational and Structural Biotechnology Journal **20**, 1036–1043 (2022). https://doi.org/10.1016/j.csbj.2022.02.005, https://www.sciencedirect.com/science/article/pii/S2001037022000423
22. Zhang, K., Zhuang, X.: CycleMix: A Holistic Strategy for Medical Image Segmentation from Scribble Supervision. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11646–11655. IEEE, New Orleans, LA, USA (Jun 2022). https://doi.org/10.1109/CVPR52688.2022.01136, https://ieeexplore.ieee.org/document/9879481/