A Transductive Few-Shot Learning Approach
for Classification of Digital Histopathological Slides from Liver Cancer.

Abdelbaki GUIR, Kaiyuan XU

M2 ATSI
Paris-Saclay University

Supervisors : Pr.Emilie Chouzenoux, Dr.Quentin VANDERBECQ.

Table of Contents

Table of Contents

1.1 Whole Slide Images (WSIs)

$\rightsquigarrow$ Digitize histology slides into high resolution images

$\rightsquigarrow$ Analyze using AI assistance systems, like conventional neural networks (CNNs), which depends on the availability of extensive annotated training data

$\rightsquigarrow$ Histopathology (v.s.natural images) : the labeling process, i.e. accurate annotations of gigapixel-sized images $\rightarrow$ time-consuming $\rightarrow$ datasets are limited in size $\Rightarrow$ data scarcity

$\rightsquigarrow$ Histopathology : non-uniform class distribution may produce biased results and compromise the model performance $\Rightarrow$ class imbalance

1.2 Classify local tissues into five classes (i.e. K = 5)

Class	NT	RE	AM	VE	AN
Percentage	26%	14%	8%	12%	40%

Table: Data distribution per class.

Non-Tumor Liver (NT), Hemorragic tissue (RE), Tumor tissue with macro-trabecular architecture (AM), Tumor tissue with Vessels Encapsulating Tumor Clusters architecture (VE), Conventional trabecular architecture (AN)

1.3 Few-shot learning (FSL)

Few-shot learning (FSL) focuses on training models to generalize from a **small** set of labeled examples.

- **Support set:** Few labeled examples for training. ($\mathbb{S} \subset \{1, \ldots, N\}$)
- **Query set:** Unlabeled examples to classify. ($\mathbb{Q} \subset \{1, \ldots, N\} \setminus \mathbb{S}$)
- **Inductive FSL:** Predicts each query sample independently of others.
  $\Rightarrow$ treats each data sample independently
- **Transductive FSL:** Uses the statistics of all unlabeled query samples collectively for improved accuracy.
  $\Rightarrow$ makes predictions on a set of samples collectively.
  $\Rightarrow$ dealing with localized regions in medical imaging ✓ (leverage `homogeneity + spatial coherence` across multiple patches)

1.4 Main contributions

Introduce a novel transductive few-shot learning approach for histopathological image classification :

1. apply a sliding window technique to WSIs $\Rightarrow$ transductive few-shot learning
2. an optimization-based method (Minimization problem)

$$\begin{aligned} \min_{U,W} \quad & f(U,W) + g(U) + \lambda h(U), \\ \text{subject to} \quad & (\forall n \in \mathbb{Q}) \quad u_n \in \Delta_K, \\ & (\forall n \in \mathbb{S}) \quad u_n = y_n, \end{aligned}$$

3. tests on the most frequent liver cancer ✓

Table of Contents

2.1 Problem formulation 1/3

$\rightsquigarrow$ The pre-trained network encoder Φ (feature extraction) :

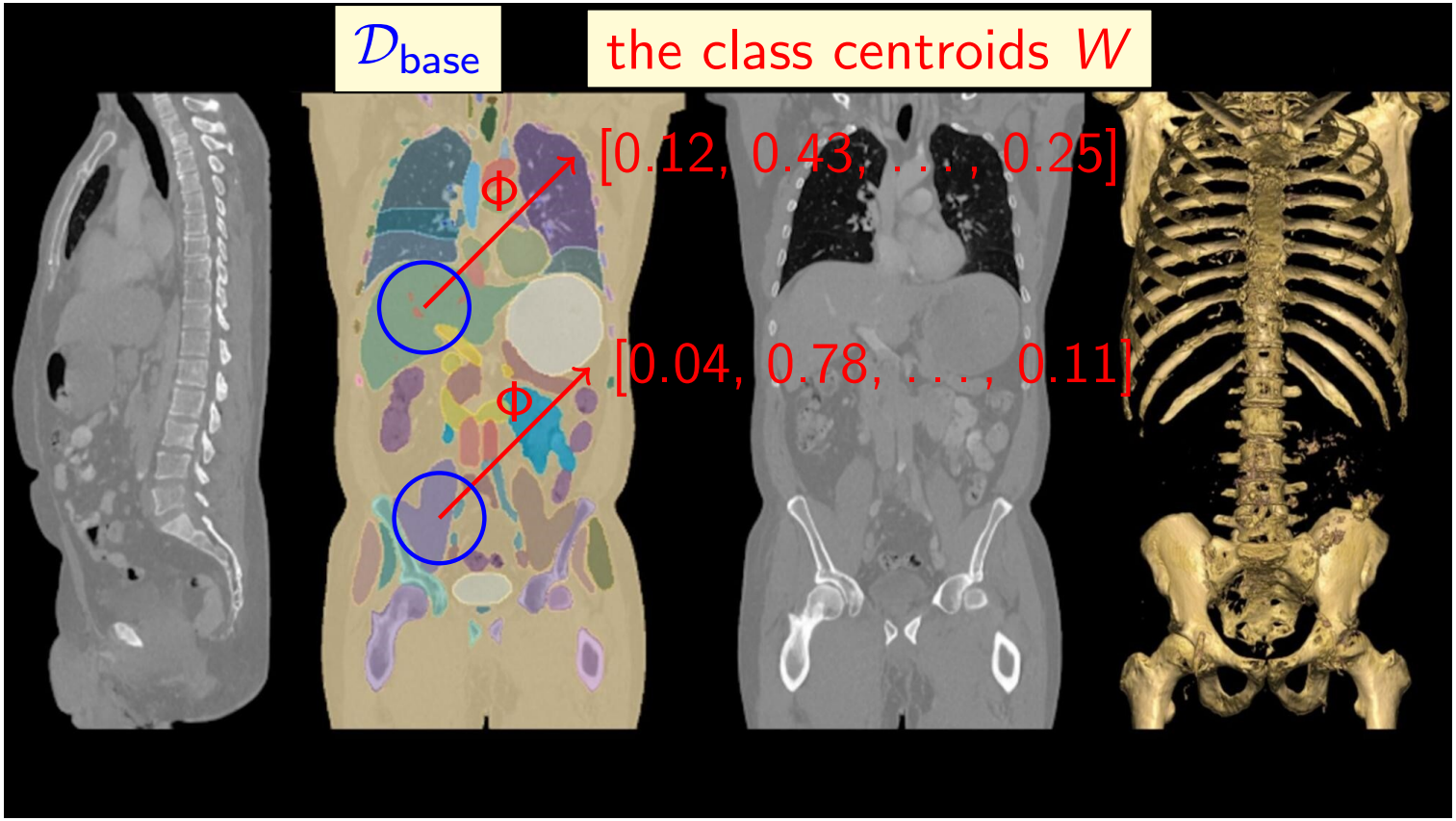


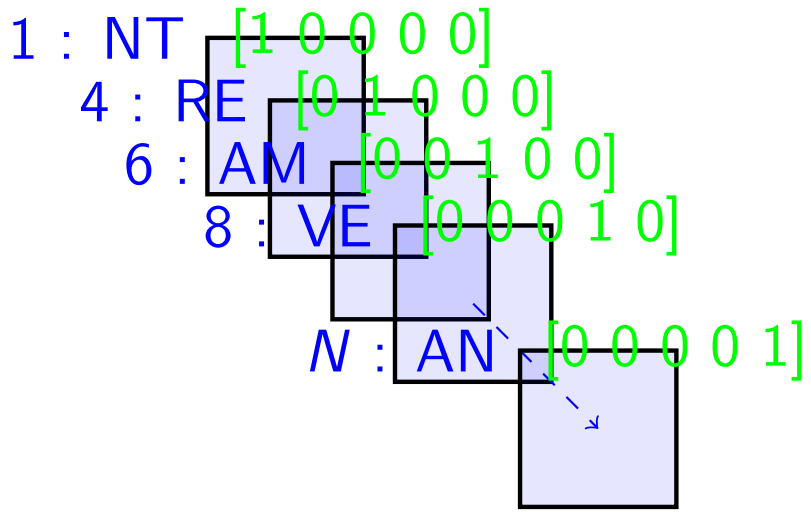
Figure: Encoder Φ : WSIs (of organs : liver, kidney, ...) $\rightarrow$ tensor .[1]

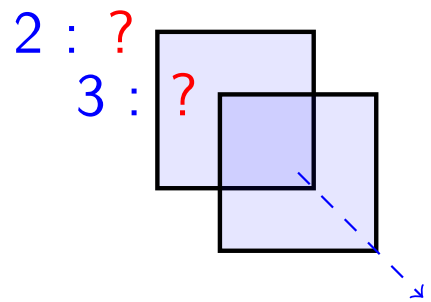
[1] https://developer.nvidia.com/blog/accelerate-medical-imaging-ai-operations-with-databricks-pixels-2-0-and-monai/

2.1 Problem formulation 2/3

$\mathbb{S} \subset \{1, \ldots, N\}$

$\mathbb{Q} \subset \{1, \ldots, N\} \setminus \mathbb{S}$





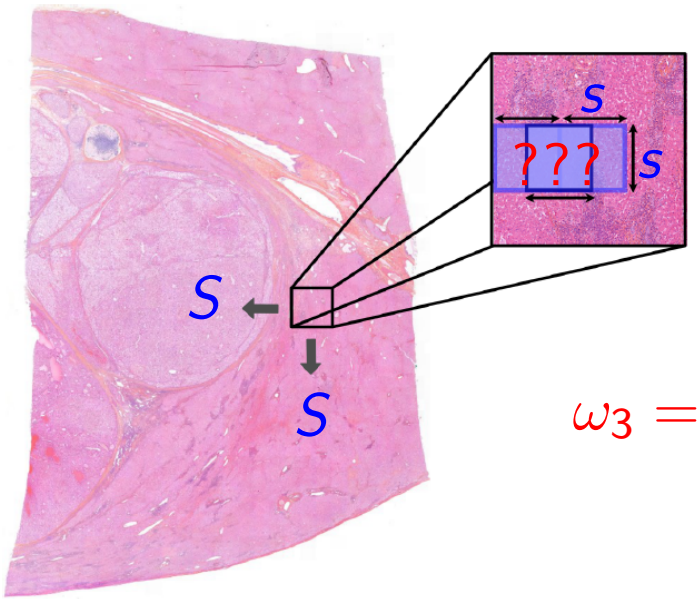
$(x_n)_{n \in \mathbb{S}}$: the feature samples

$(x_n)_{n \in \mathbb{Q}}$: unlabeled samples

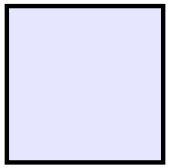
$(y_n)_{n \in \mathbb{S}}$: one-hot-encoded labels

? : put $(z_n = \Phi(x_n))_{1 \leq n \leq N}$ into the few-shot classifier

2.1 Problem formulation 3/3



2 : ?



$u_2 = [0\ 0\ 1\ 0\ 0] \Rightarrow$ AM, i.e. k $= 3$

$\omega_3 = [0.12, 0.43, ..., 0.25] \Rightarrow$ AM, i.e. k $= 3$

Figure: Scanning of the slide with a sliding window. cf. p.3 [2]

$\rightsquigarrow$ $\mathbf{U} = (u_n)_{1 \leq n \leq |\mathbb{Q}|} \in (\Delta_K)^{|\mathbb{Q}|}$: one-hot-encoded assignments

$\rightsquigarrow$ $\mathbf{W} = (w_k)_{1 \leq k \leq K} \in (\mathbb{R}^d)^K$: the class centroids

[2] https://arxiv.org/abs/2311.17740

2.2 Minimization problem

Problem formulation

$$\begin{aligned} \min_{U,W} \quad & f(U,W) + g(U) + \lambda h(U), \\ \text{subject to} \quad & (\forall n \in \mathbb{Q}) \quad u_n \in \Delta_K, \\ & (\forall n \in \mathbb{S}) \quad u_n = y_n, \end{aligned} \tag{1}$$

Terms:

1. **Data-fidelity term ($f(U, W)$):** Assumption that the data follows a `multivariate Gaussian distribution` and integrating supervision from the `support set`.

$$f(U, W) = \frac{1}{2} \sum_{k=1}^{K} \sum_{n=1}^{N} u_{n,k} (w_k - z_n)^\top \hat{S}_k (w_k - z_n) - \frac{1}{2} \sum_{k=1}^{K} \sum_{n=1}^{N} u_{n,k} \ln \det(\hat{S}_k)$$

2. **Entropic barrier on the assignments ($g(U)$):** Facilitating `closed-form` updates in the forthcoming algorithm.

$$g(U) = \sum_{k=1}^{K} \sum_{n \in \mathbb{Q}} u_{n,k} \ln u_{n,k}.$$

3. **Penalty function ($h(U)$):** Encouraging a minimal number of classes to be predicted within the window.

$$h(U) = -\sum_{k=1}^{K} \pi_k \ln(\pi_k)$$

2.3 Algorithm

Algorithm 1: PADDLE-Cov

Initialize $\boldsymbol{W}^{(0)}$ as the means computed on the support set and for all $k \in \{1, \ldots, K\}$,

$$\pi_k^{(0)} = \frac{1}{|\mathbb{Q}|} \sum_{n \in \mathbb{Q}} u_{n,k}^{(0)}.$$

for $\ell = 1, 2, \ldots,$ **do**

$$\boldsymbol{u}_n^{(\ell)} = \operatorname{softmax}\left(\left(-\frac{1}{2}(\boldsymbol{w}_k - \boldsymbol{z}_n)^\top \hat{\boldsymbol{S}}_{\boldsymbol{k}}(\boldsymbol{w}_k - \boldsymbol{z}_n) + \tfrac{1}{2}\ln\det(\hat{\boldsymbol{S}}_k) + \frac{\lambda}{|\mathbb{Q}|}\ln \pi_k^{(\ell)}\right)_k\right), \ \forall n \in \mathbb{Q},$$

$$\boldsymbol{w}_k^{(\ell+1)} = \frac{\sum_{n=1}^N u_{n,k}^{(\ell+1)} \boldsymbol{z}_n}{\sum_{n=1}^N u_{n,k}^{(\ell+1)}}, \ \forall k \in \{1, \ldots, K\},$$

$$\pi_k^{(\ell+1)} = \frac{1}{|\mathbb{Q}|} \sum_{n \in \mathbb{Q}} u_{n,k}^{(\ell+1)}, \quad \forall k \in \{1, \ldots, K\}.$$

Figure: PADDLE-Cov. cf. p.3 [3]

[3] https://arxiv.org/abs/2412.16739v1

Table of Contents

3.1 Experimental setting

- $S = 5184$, $s = 1728 \rightarrow 512 \Rightarrow |\mathbb{Q}| = 25$
- $\Rightarrow |\mathbb{S}| = 28$
- $\lambda = 1250$

3.2 Results

Method	**Accuracy (%)**	**F1-score (%)**
SimpleShot	48.9	46.4
Baseline	74.4	72.0
α-TIM	56.0	56.9
PADDLE	51.0	48.9
PADDLE-Cov ($\lambda = 0$)	77.3	73.8
PADDLE-Cov	**79.3**	**75.5**

Table: Evaluation of our approach against other few-shot methods for histopathological patch classification regarding accuracy and F1-score. cf. p.4 [4]

[4] https://arxiv.org/abs/2412.16739v1

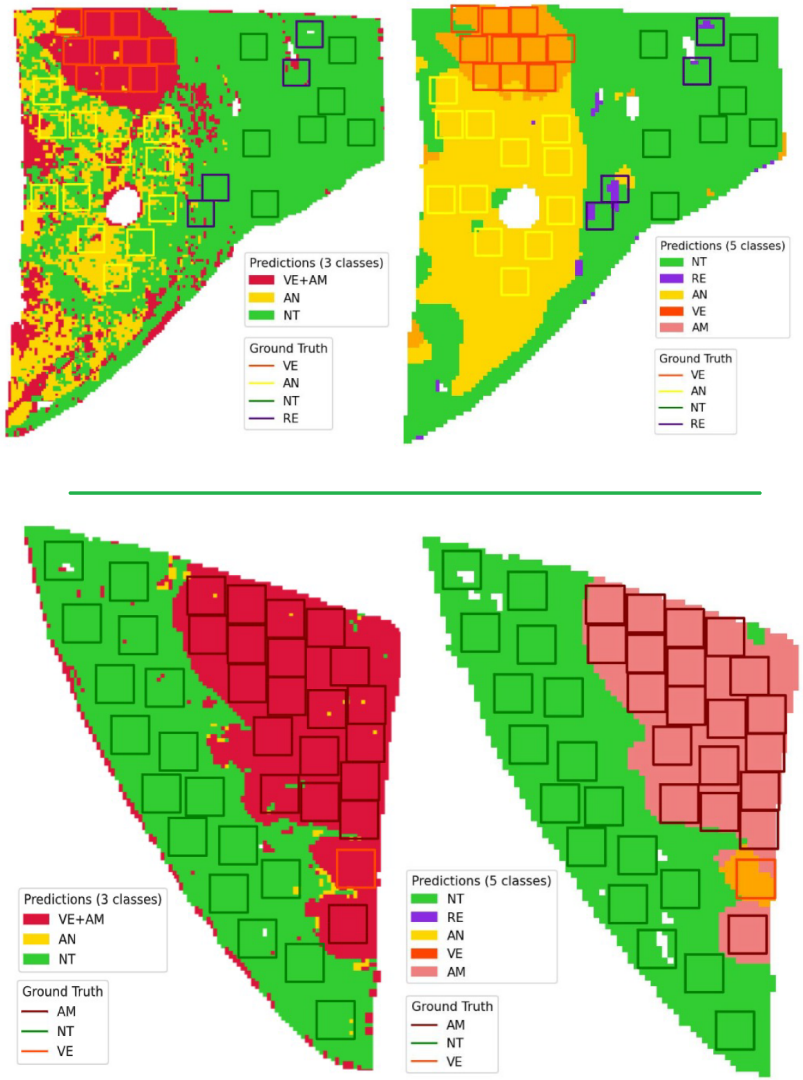


Figure: (left) The 3-class fully supervised model predictions. (right) The few-shot 5-class model predictions. cf. p.4 [5]

[5] https://arxiv.org/abs/2412.16739v1

Table of Contents

Questions/Answers

Algorithm 1: PADDLE-Cov

Initialize $\boldsymbol{W}^{(0)}$ as the means computed on the support set and for all $k \in \{1, \ldots, K\}$,

$$\pi_k^{(0)} = \frac{1}{|\mathbb{Q}|} \sum_{n \in \mathbb{Q}} u_{n,k}^{(0)}.$$

for $\ell = 1, 2, \ldots,$ **do**

$$\boldsymbol{u}_n^{(\ell)} = \operatorname{softmax}\left(\left(-\frac{1}{2} (\boldsymbol{w}_k - \boldsymbol{z}_n)^\top \hat{\boldsymbol{S}}_{\boldsymbol{k}} (\boldsymbol{w}_k - \boldsymbol{z}_n) + \tfrac{1}{2} \ln \det(\hat{\boldsymbol{S}}_k) + \tfrac{\lambda}{|\mathbb{Q}|} \ln \pi_k^{(\ell)} \right)_k \right), \ \forall n \in \mathbb{Q},$$

$$\boldsymbol{w}_k^{(\ell+1)} = \frac{\sum_{n=1}^N u_{n,k}^{(\ell+1)} \boldsymbol{z}_n}{\sum_{n=1}^N u_{n,k}^{(\ell+1)}}, \ \forall k \in \{1, \ldots, K\},$$

$$\pi_k^{(\ell+1)} = \frac{1}{|\mathbb{Q}|} \sum_{n \in \mathbb{Q}} u_{n,k}^{(\ell+1)}, \quad \forall k \in \{1, \ldots, K\}.$$

$\hat{S}_k$ is a symmetric positive matrix corresponding to a sparse approximation of inverse of the empirical covariance matrix of class k, computed from the support set with a Graphical Lasso approach [6].

$$\omega_k^{(\ell+1)} = \frac{\sum_{n=1}^{N} u_{n,k}^{(\ell+1)} \mathbf{z}_n}{\sum_{n=1}^{N} u_{n,k}^{(\ell+1)}}$$

$$\Sigma_k^{(\ell+1)} \stackrel{?}{=} \frac{\sum_{n=1}^{N} u_{n,k}^{(\ell+1)} (\mathbf{z}_n - \omega_k^{(\ell+1)})(\mathbf{z}_n - \omega_k^{(\ell+1)})^T}{\sum_{n=1}^{N} u_{n,k}^{(\ell+1)}}$$

$u_{n,k}$: Soft-Assignment vectors (Probability that the n-th sample belongs to k-th class)

$\mathbf{U} = (u_n)_{1 \leq n \leq |\mathbb{Q}|} \in (\Delta_K)^{|\mathbb{Q}|}$: one-hot-encoded assignments

[6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," Biostatistics, vol. 9, no. 3, pp. 432–441, 2008.

Thanks for your attention.