Weakly-supervised positional contrastive learning: application to cirrhosis classification

Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, Pietro Gori, Isabelle Bloch

Weakly-supervised positional contrastive learning: application to cirrhosis classification

Emma Sarfati[1,2], Alexandre Bône[1], Marc-Michel Rohé[1], Pietro Gori[2], and Isabelle Bloch[2,3]

[1] Guerbet Research, Villepinte, France
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[3] Sorbonne Université, CNRS, LIP6, Paris, France

Abstract. Large medical imaging datasets can be cheaply and quickly annotated with low-confidence, weak labels (*e.g.*, radiological scores). Access to high-confidence labels, such as histology-based diagnoses, is rare and costly. Pretraining strategies, like contrastive learning (CL) methods, can leverage unlabeled or weakly-annotated datasets. These methods typically require large batch sizes, which poses a difficulty in the case of large 3D images at full resolution, due to limited GPU memory. Nevertheless, volumetric positional information about the spatial context of each 2D slice can be very important for some medical applications. In this work, we propose an efficient weakly-supervised positional (WSP) contrastive learning strategy where we integrate both the spatial context of each 2D slice and a weak label via a generic kernel-based loss function. We illustrate our method on cirrhosis prediction using a large volume of weakly-labeled images, namely radiological low-confidence annotations, and small strongly-labeled (*i.e.*, high-confidence) datasets. The proposed model improves the classification AUC by 5% with respect to a baseline model on our internal dataset, and by 26% on the public LIHC dataset from the Cancer Genome Atlas. The code is available at: `https://github.com/Guerbet-AI/wsp-contrastive`.

Keywords: Weakly-supervised learning, Contrastive learning, CT, Cirrhosis prediction, Liver.

1 Introduction

In the medical domain, obtaining a large amount of high-confidence labels, such as histopathological diagnoses, is arduous due to the cost and required technicality. It is however possible to obtain lower confidence assessments for a large amount of images, either by a clinical questioning, or directly by a radiological diagnosis. To take advantage of large volumes of unlabeled or weakly-labeled images, pre-training encoders with self-supervised methods showed promising results in deep learning for medical imaging [1,4,21,27,28,29]. In particular, contrastive learning (CL) is a self-supervised method that learns a mapping of the input images to a representation space where similar (positive) samples are moved

closer and different (negative) samples are pushed far apart. Weak discrete labels can be integrated into contrastive learning by, for instance, considering as positives only the samples having the same label, as in [13], or by directly weighting unsupervised contrastive and supervised cross entropy loss functions, as in [19]. In this work, we focus on the scenario where radiological meta-data (thus, low-confidence labels) are available for a large amount of images, whereas high-confidence labels, obtained by histological analysis, are scarce.

Naive extensions of contrastive learning methods, such as [5,10,11], from 2D to 3D images may be difficult due to limited GPU memory and therefore small batch size. A usual solution consists in using patch-based methods [8,23]. However, these methods pose two difficulties: they reduce the spatial context (limited by the size of the patch), and they require similar spatial resolution across images. This is rarely the case for abdominal CT/MRI acquisitions, which are typically strongly anisotropic and with variable resolutions. Alternatively, depth position of each 2D slice, within its corresponding volume, can be integrated in the analysis. For instance, in [4], the authors proposed to integrate depth in the sampling strategy for the batch creation. Likewise, in [26], the authors proposed to define as similar only 2D slices that have a *small* depth difference, using a normalized depth coordinate $d \in [0, 1]$. These works implicitly assume a certain threshold on depth to define positive and negative samples, which may be difficult to define and may be different among applications and datasets. Differently, inspired by [2,8], here we propose to use a degree of "positiveness" between samples by defining a kernel function w on depth positions. This allows us to consider volumetric depth information during pre-training *and* to use large batch sizes. Furthermore, we also propose to *simultaneously* leverage weak discrete attributes during pre-training by using a novel and efficient contrastive learning composite kernel loss function, denoting our global method Weakly-Supervised Positional (WSP).

We apply our method to the classification of histology-proven liver cirrhosis, with a large volume of (weakly) radiologically-annotated CT-scans and a small amount of histopathologically-confirmed cirrhosis diagnosis. We compare the proposed approach to existing self-supervised methods.

2 Method

Let x_t be an input 2D image, usually called *anchor*, extracted from a 3D volume, y_t a corresponding discrete weak variable and d_t a related continuous variable. In this paper, y_t refers to a weak radiological annotation and d_t corresponds to the normalized depth position of the 2D image within its corresponding 3D volume: if V_{max} corresponds to the maximal depth-coordinate of a volume V, we compute $d_t = \frac{p_t}{V_{max}}$ with $p_t \in [0, V_{max}]$ being the original depth coordinate. Let x_j^- and x_i^+ be two semantically different (negative) and similar (positive) images with respect to x_t, respectively.

The definition of similarity is crucial in CL and is the main difference between existing methods. For instance, in unsupervised CL, methods such as

SimCLR [5,6] choose as positive samples random augmentations of the anchor $x_i^+ = t(x_t)$, where $t \sim \mathcal{T}$ is a random transformation chosen among a user-selected family $\mathcal{T}$. Negative images x_j^- are all other (transformed) images present in the batch.

Once x_j^- and x_i^+ are defined, the goal of CL is to compute a mapping function $f_\theta : \mathcal{X} \to \mathbb{S}^d$, where $\mathcal{X}$ is the set of images and $\mathbb{S}^d$ the representation space, so that similar samples are mapped closer in the representation space than dissimilar samples. Mathematically, this can be defined as looking for a f_θ that satisfies the condition:

$$s_{tj}^- - s_{ti}^+ \leq 0 \quad \forall t, j, i \tag{1}$$

where $s_{tj}^- = sim(f_\theta(x_t), f_\theta(x_j^-))$ and $s_{ti}^+ = sim(f_\theta(x_t), f_\theta(x_i^+))$, with sim a similarity function defined here as $sim(a, b) = \frac{a^T b}{\tau}$ with $\tau > 0$.

In the presence of discrete labels y, the definition of negative (x_j^-) and positive (x_i^+) samples may change. For instance, in SupCon [13], the authors define as positives all images with the same discrete label y. However, when working with continuous labels d, one cannot use the same strategy since all images are somehow positive and negative at the same time. A possible solution [26] would be to define a threshold γ on the distance between labels (*e.g.*, d_a, d_b) so that, if the distance is smaller than γ (*i.e.*, $||d_a - d_b||_2 < \gamma$), the samples (*e.g.*, x_a and x_b) are considered as positives. However, this requires a user-defined hyper-parameter γ, which could be hard to find in practice. A more efficient solution, as proposed in [8], is to define a degree of "positiveness" between samples using a normalized kernel function $w_\sigma(d, d_i) = K_\sigma(d - d_i)$, where K_σ is, for instance, a Gaussian kernel, with user defined hyper-parameter σ and $0 \leq w_\sigma \leq 1$. It is interesting to notice that, for discrete labels, one could also define a kernel as: $w_\delta(y, y_i) = \delta(y - y_i)$, δ being the Dirac function, retrieving exactly SupCon [13].

In this work, we propose to leverage both continuous d and discrete y labels, by combining (here by multiplying) the previously defined kernels, w_σ and w_δ, into a composite kernel loss function. In this way, samples will be considered as similar (positive) only if they have a *composite* degree of "positiveness" greater than zero, namely both kernels have a value greater (or different) than 0 ($w_\sigma > 0$ and $w_\delta \neq 0$). An example of resulting representation space is shown in Figure 1. This constraint can be defined by slightly modifying the condition introduced in Equation 1, as:

$$\underbrace{w_\delta(y_t, y_i) \cdot w_\sigma(d_t, d_i)}_{\text{composite kernel } w_{ti}}(s_{tj} - s_{ti}) \leq 0 \quad \forall t, i, j \neq i \tag{2}$$

where the indices t, i, j traverse all N images in the batch since there are no "hard" positive or negative samples, as in SimCLR or SupCon, but all images are considered as positive and negative at the same time. As commonly done in CL [3], this condition can be transformed into an optimization problem using

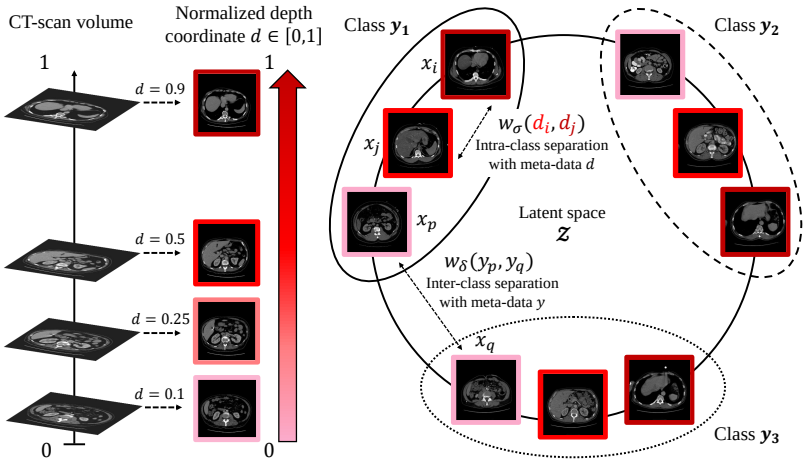


Fig. 1: Example of representation space constructed by our loss function, leveraging both continuous depth coordinate d and discrete label y (*i.e.*, radiological diagnosis y_{radio}). Samples from different radiological classes are well separated and, at the same time, samples are ordered within each class based on their depth coordinate d.

the max operator and its smooth approximation *LogSumExp*:

$$\begin{aligned}\arg\min_{f_\theta} \sum_{t,i} \max(0, w_{ti}\{s_{tj} - s_{ti}\}_{\substack{j=1\\ j\neq i}}^{N}) &= \arg\min_{f_\theta} \sum_{t,i} w_{ti} \max(0, \{s_{tj} - s_{ti}\}_{\substack{j=1\\ j\neq i}}^{N}) \\ &\approx \arg\min_{f_\theta} \left(-\sum_{t,i} w_{ti} \log \left(\frac{\exp(s_{ti})}{\sum_{j\neq i}^{N} \exp(s_{tj})} \right) \right)\end{aligned} \tag{3}$$

By defining $P(t) = \{i : y_i = y_t\}$ as the set of indices of images x_i in the batch with the same discrete label y_i as the anchor x_t, we can rewrite our final loss function as:

$$\mathcal{L}_{WSP} = -\sum_{t=1}^{N} \sum_{i \in P(t)} w_\sigma(d_t, d_i) \log \left(\frac{\exp(s_{ti})}{\sum_{j\neq i}^{N} \exp(s_{tj})} \right) \tag{4}$$

where $w_\sigma(d_t, d_i)$ is normalized over $i \in P(t)$. In practice, it is rather easy to find a good value of σ, as the proposed kernel method is quite robust to its variation. A robustness study is available in the supplementary material. For the experiments, we fix $\sigma = 0.1$.

3 Experiments

We compare the proposed method with different contrastive and non-contrastive methods, that either use no meta-data (SimCLR [5], BYOL [10]), or leverage

only discrete labels (SupCon [13]), or continuous labels (depth-Aware [8]). The proposed method is the only one that takes simultaneously into account both discrete and continuous labels. In all experiments, we work with 2D slices rather than 3D volumes due to the anisotropy of abdominal CT-scans in the depth direction and the limited spatial context or resolution obtained with 3D patch-based or downsampling methods, respectively, which strongly impacts the cirrhosis diagnosis that is notably based on the contours irregularity. Moreover, the large batch sizes necessary in contrastive learning can not be handled in 3D due to a limited GPU memory.

3.1 Datasets

Three datasets of abdominal CT images are used in this study. One dataset is used for contrastive pretraining, and the other two for evaluation. All images have a 512x512 size, and we clip the intensity values between -100 and 400.

$\boldsymbol{\mathcal{D}_{radio}}$. First, $\mathcal{D}_{radio}$ contains 2,799 CT-scans of patients in portal venous phase with a radiological (weak) annotation, *i.e.* realized by a radiologist, indicating four different stages of cirrhosis: no cirrhosis, mild cirrhosis, moderate cirrhosis and severe cirrhosis (y_{radio}). The respective numbers are 1880, 385, 415 and 119. y_{radio} is used as the discrete label y during pre-training.

$\boldsymbol{\mathcal{D}^1_{histo}}$. It contains 106 CT-scans from different patients in portal venous phase, with an identified histopathological status (METAVIR score) obtained by a histological analysis, designated as y^1_{histo}. It corresponds to absent fibrosis (F0), mild fibrosis (F1), significant fibrosis (F2), severe fibrosis (F3) and cirrhosis (F4). This score is then binarized to indicate the absence or presence of advanced fibrosis [14]: F0/F1/F2 (N=28) vs. F3/F4 (N=78).

$\boldsymbol{\mathcal{D}^2_{histo}}$. This is the public LIHC dataset from the Cancer Genome Atlas [9], which presents a histological score, the Ishak score, designated as y^2_{histo}, that differs from the METAVIR score present in $\mathcal{D}^1_{histo}$. This score is also distributed through five labels: No Fibrosis, Portal Fibrosis, Fibrous Speta, Nodular Formation and Incomplete Cirrhosis and Established Cirrhosis. Similarly to the METAVIR score in $\mathcal{D}^1_{histo}$, we also binarize the Ishak score, as proposed in [16,20], which results in two cohorts of 34 healthy and 15 pathological patients.

In all datasets, we select the slices based on the liver segmentation of the patients. To gain in precision, we keep the top 70% most central slices with respect to liver segmentation maps obtained manually in $\mathcal{D}_{radio}$, and automatically for $\mathcal{D}^1_{histo}$ and $\mathcal{D}^2_{histo}$ using a U-Net architecture pretrained on $\mathcal{D}_{radio}$ [18]. For the latter pretraining dataset, it presents an average slice spacing of 3.23mm with a standard deviation of 1.29mm. For the x and y axis, the dimension is 0.79mm per voxel on average, with a standard deviation of 0.10mm.

3.2 Architecture and optimization.

Backbones. We propose to work with two different backbones in this paper: TinyNet and ResNet-18 [12]. TinyNet is a small encoder with 1.1M parameters,

inspired by [24], with five convolutional layers, a representation space (for downstream tasks) of size 256 and a latent space (after a projection head of two dense layers) of size 64. In comparison, ResNet-18 has 11.2M parameters, a representation space of dimension 512 and a latent space of dimension 128. More details and an illustration of TinyNet are available in the supplementary material, as well as a full illustration of the algorithm flow.

Data augmentation, sampling and optimization. CL methods [5,10,11] require strong data augmentations on input images, in order to strengthen the association between positive samples [22]. In our work, we leverage three types of augmentations: rotations, crops and flips. Data augmentations are computed on the GPU, using the Kornia library [17]. During inference, we remove the augmentation module to only keep the original input images.

For sampling, inspired by [4], we propose a strategy well-adapted for contrastive learning in 2D medical imaging. We first sample N patients, where N is the batch size, in a balanced way with respect to the radiological/histological classes; namely, we roughly have the same number of subjects per class. Then, we randomly select only one slice per subject. In this way, we maximize the slice heterogeneity within each batch. We use the same sampling strategy also for classification baselines. For $\mathcal{D}^2_{histo}$, which has fewer patients than the batch size, we use a balanced sampling strategy with respect to the radiological/histological classes with no obligation of one slice per patient in the batch. As we work with 2D slices rather than 3D volumes, we compute the average probability per patient of having the pathology. The evaluation results presented later are based on the patient-level aggregated prediction.

Finally, we run our experiments on a Tesla V100 with 16GB of RAM and a 6 CPU cores, and we used the PyTorch-Lightning library to implement our models. All models share the same data augmentation module, with a batch size of $B = 64$ and a fixed number of epochs $n_{epochs} = 200$. For all experiments, we fix a learning rate (LR) of $\alpha = 10^{-4}$ and a weight decay of $\lambda = 10^{-4}$. We add a cosine decay learning rate scheduler [15] to prevent over-fitting. For BYOL, we initialize the moving average decay at 0.996.

Evaluation protocol. We first pretrain the backbone networks on $\mathcal{D}_{radio}$ using all previously listed contrastive and non-contrastive methods. Then, we train a regularized logistic regression on the frozen representations of the datasets $\mathcal{D}^1_{histo}$ and $\mathcal{D}^2_{histo}$. We use a stratified 5-fold cross-validation. As a baseline, we train a classification algorithm from scratch (supervised) for each dataset, $\mathcal{D}^1_{histo}$ and $\mathcal{D}^2_{histo}$, using both backbone encoders and the same 5-fold cross-validation strategy. We also train a regularized logistic regression on representations obtained with a random initialization as a second baseline (random). Finally, we report the cross-validated results for each model on the aggregated dataset $\mathcal{D}^{1+2}_{histo} = \mathcal{D}^1_{histo} + \mathcal{D}^2_{histo}$.

Table 1: Resulting 5-fold cross-validation AUCs. For each encoder, best results are in **bold**, second top results are underlined. * = We use the pretrained weights from ImageNet with ResNet-18 and run a logistic regression on the frozen representations.

Backbone	Pretraining method	Weak labels	Depth pos.	$\mathcal{D}^1_{histo}$ (N=106)	$\mathcal{D}^2_{histo}$ (N=49)	$\mathcal{D}^{1+2}_{histo}$ (N=155)
TinyNet	Supervised	✗	✗	0.79 (±0.05)	0.65 (±0.25)	0.71 (±0.04)
	None (random)	✗	✗	0.64 (±0.10)	0.75 (±0.13)	0.73 (±0.06)
	SimCLR	✗	✗	0.75 (±0.08)	0.88 (±0.16)	0.76 (±0.11)
	BYOL	✗	✗	0.75 (±0.09)	**0.95 (±0.07)**	0.77 (±0.08)
	SupCon	✓	✗	0.76 (±0.09)	0.93 (±0.07)	0.72 (±0.06)
	depth-Aware	✗	✓	0.80 (±0.13)	0.81 (±0.08)	0.77 (±0.08)
	Ours	✓	✓	**0.84 (±0.12)**	0.91 (±0.11)	**0.79 (±0.11)**
ResNet-18	Supervised	✗	✗	0.77 (±0.10)	0.56 (±0.29)	0.72 (±0.08)
	None (random)	✗	✗	0.69 (±0.19)	0.73 (±0.12)	0.68 (±0.09)
	ImageNet*	✗	✗	0.72 (±0.17)	0.76 (±0.04)	0.66 (±0.10)
	SimCLR	✗	✗	0.79 (±0.09)	0.82 (±0.14)	0.79 (±0.08)
	BYOL	✗	✗	0.78 (±0.09)	0.77 (±0.11)	0.78 (±0.08)
	SupCon	✓	✗	0.69 (±0.07)	0.69 (±0.13)	0.76 (±0.12)
	depth-Aware	✗	✓	0.83 (±0.07)	0.82 (±0.11)	0.80 (±0.07)
	Ours	✓	✓	**0.84 (±0.07)**	**0.85 (±0.10)**	**0.84 (±0.07)**

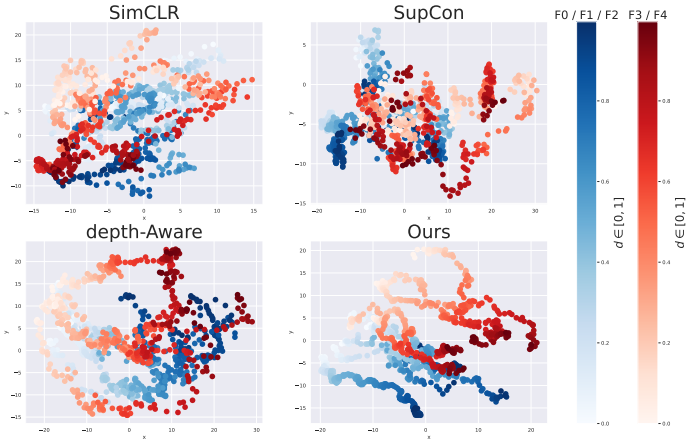


Fig. 2: Projections of the ResNet-18 representation vectors of 10 randomly selected subjects of $\mathcal{D}^1_{histo}$ onto the first two modes of a PCA. Each dot represents a 2D slice. Color gradient refers to different depth positions. Red = cirrhotic cases. Blue = healthy subjects.

4 Results and Discussion

We present in Table 1 the results of all our experiments. For each of them, we report whether the pretraining method integrates the weak label meta-data, the depth spatial encoding, or both, which is the core of our method. First, we can notice that our method outperforms all other pretraining methods in $\mathcal{D}^1_{histo}$ and $\mathcal{D}^{1+2}_{histo}$, which are the two datasets with more patients. For the latter, the proposed method surpasses the second best pretraining method, depth-Aware, by 4%. For $\mathcal{D}^1_{histo}$, it can be noticed that WSP (ours) provides the best AUC score whatever the backbone used. For the second dataset $\mathcal{D}^2_{histo}$, our method is on par with BYOL and SupCon when using a small encoder and outperforms the other methods when using a larger backbone.

To illustrate the impact of the proposed method, we report in Figure 2 the projections of the ResNet-18 representation vectors of 10 randomly selected subjects of $\mathcal{D}^1_{histo}$ onto the first two modes of a PCA. It can be noticed that the representation space of our method is the only one where the diagnostic label (not available during pretraining) and the depth position are correctly integrated. Indeed, there is a clear separation between slices of different classes (healthy at the bottom and cirrhotic cases at the top) and at the same time it seems that the depth position has been encoded in the x-axis, from left to right. SupCon performs well on the training set of $\mathcal{D}_{radio}$ (figure available in the supplementary material), as well as $\mathcal{D}^2_{histo}$ with TinyNet, but it poorly generalizes to $\mathcal{D}^1_{histo}$ and $\mathcal{D}^{1+2}_{histo}$. The method depth-Aware manages to correctly encode the depth position but not the diagnostic class label.

To assess the clinical performance of the pretraining methods, we also compute the balanced accuracy scores (bACC) of the trained classifiers, which is compared in Table 2 to the bACC achieved by radiologists who were asked to visually assess the presence or absence of cirrhosis for the N=106 cases of $\mathcal{D}^1_{histo}$.

Table 2: Comparison of the pretraining methods with a binary radiological annotation for cirrhosis on $\mathcal{D}^1_{histo}$. Best results are in **bold**, second top results are underlined.

Pretraining method	**bACC models**	**bACC radiologists**
Supervised	0.78 (±0.04)	
None (random)	0.71 (±0.13)	
ImageNet	0.74 (±0.13)	
SimCLR	0.78 (±0.08)	
BYOL	0.77 (±0.04)	0.82
SupCon	0.77 (±0.10)	
depth-Aware	<u>0.84 (±0.04)</u>	
Ours	**0.85 (±0.09)**	

The reported bACC values correspond to the best scores among those obtained with Tiny and ResNet encoders. Radiologists achieved a bACC of 82% with respect to the histological reference. The two best-performing methods surpassed this score: depth-Aware and the proposed WSP approach, improving respectively the radiologists score by 2% and 3%, suggesting that including 3D information (depth) at the pretraining phase was beneficial.

5 Conclusion

In this work, we proposed a novel kernel-based contrastive learning method that leverages both continuous and discrete meta-data for pretraining. We tested it on a challenging clinical application, cirrhosis prediction, using three different datasets, including the LIHC public dataset. To the best of our knowledge, this is the first time that a pretraining strategy combining different kinds of meta-data has been proposed for such application. Our results were compared to other state-of-the-art CL methods well-adapted for cirrhosis prediction. The pretraining methods were also compared visually, using a 2D projection of the representation vectors onto the first two PCA modes. Results showed that our method has an organization in the representation space that is in line with the proposed theory, which may explain its higher performances in the experiments. As future work, it would be interesting to adapt our kernel method to non-contrastive methods, such as SimSIAM [7], BYOL [10] or Barlow Twins [25], that need smaller batch sizes and have shown greater perfomances in computer vision tasks. In terms of application, our method could be easily translated to other medical problems, such as pancreas cancer prediction using the presence of intrapancreatic fat, diabetes mellitus or obesity as discrete meta-labels.

Compliance with ethical standards. This research study was conducted retrospectively using human data collected from various medical centers, whose Ethics Committees granted their approval. Data was de-identified and processed according to all applicable privacy laws and the Declaration of Helsinki.

Acknowledgments. This work was supported by Région Ile-de-France (ChoTherIA project) and ANRT (CIFRE #2021/1735).

References

1. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V., Norouzi, M.: Big self-supervised models advance medical image classification. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3458–3468 (2021)
2. Barbano, C.A., Dufumier, B., Duchesnay, E., Grangetto, M., Gori, P.: Contrastive learning for regression in multi-site brain age prediction. In: IEEE ISBI (2022)
3. Barbano, C.A., Dufumier, B., Tartaglione, E., Grangetto, M., Gori, P.: Unbiased Supervised Contrastive Learning. In: ICLR (2023)
4. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 12546–12558. Curran Associates, Inc. (2020)
5. Chen, T., Kornblith, S., Norouzi, M., et al.: A simple framework for contrastive learning of visual representations. In: 37th International Conference on Machine Learning (ICML) (2020)
6. Chen, T., Kornblith, S., Swersky, K., et al.: Big self-supervised models are strong semi-supervised learners. In: NeurIPS (2020)

7. Chen, X., He, K.: Exploring simple siamese representation learning. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 15745–15753 (2020)
8. Dufumier, B., Gori, P., Victor, J., et al.: Contrastive learning with continuous proxy meta-data for 3D MRI classification. In: MICCAI. pp. 58–68. Springer (2021)
9. Erickson, B.J., Kirk, S., Lee, et al.: Radiology data from the cancer genome atlas colon adenocarcinoma [TCGA-COAD] collection. (2016)
10. Grill, J.B., Strub, F., Altché, F., et al.: Bootstrap your own latent - a new approach to self-supervised learning. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 21271–21284. Curran Associates, Inc. (2020)
11. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9726–9735 (2020)
12. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)
13. Khosla, P., Teterwak, P., Wang, C., et al.: Supervised contrastive learning. Advances in Neural Information Processing Systems **33**, 18661–18673 (2020)
14. Li, Q., Yu, B., Tian, X., et al.: Deep residual nets model for staging liver fibrosis on plain CT images. International Journal of Computer Assisted Radiology and Surgery **15** (06 2020)
15. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: International Conference on Learning Representations (2017)
16. Mohamadnejad, M., Tavangar, S.M., Kosari, F., Sotoudeh, M., Khosravi, M., Geramizadeh, B., G, M., Estakhri, A., Mirnasseri, M., A, F., Zamani, F., Malekzadeh, R.: Histopathological study of chronic hepatitis B: A comparative study of ishak and METAVIR scoring systems. International Journal of Organ Transplantation Medicine **1** (11 2010)
17. Riba, E., Mishkin, D., Ponsa, D., Rublee, E., Bradski, G.: Kornia: an open source differentiable computer vision library for PyTorch. In: Winter Conference on Applications of Computer Vision (2020)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
19. Sarfati, E., Bone, A., Rohe, M.M., Gori, P., Bloch, I.: Learning to diagnose cirrhosis from radiological and histological labels with joint self and weakly-supervised pretraining strategies. In: IEEE ISBI. Cartagena de Indias, Colombia (Apr 2023)
20. Shiha, G., Zalata, K.: Ishak versus METAVIR: Terminology, convertibility and correlation with laboratory changes in chronic hepatitis C. In: Takahashi, H. (ed.) Liver Biopsy, chap. 10. IntechOpen, Rijeka (2011)
21. Taleb, A., Kirchler, M., Monti, R., Lippert, C.: Contig: Self-supervised multimodal contrastive learning for medical imaging with genetics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20908–20921 (June 2022)
22. Wang, X., Qi, G.J.: Contrastive learning with stronger augmentations. CoRR **abs/2104.07713** (2021)
23. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O.: Convolutional neural

networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis **63**, 101694 (2020)
24. Yin, Y., Yakar, D., Dierckx, R., et al.: Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model. European Radiology **31** (05 2021)
25. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning (2021)
26. Zeng, D., Wu, Y., Hu, X., Xu, X., Yuan, H., Huang, M., Zhuang, J., Hu, J., Shi, Y.: Positional contrastive learning for volumetric medical image segmentation. In: MICCAI. p. 221–230. Springer-Verlag, Berlin, Heidelberg (2021)
27. Zhang, P., Wang, F., Zheng, Y.: Self supervised deep representation learning for fine-grained body part recognition. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 578–582 (2017)
28. Zhou, Z., Sodha, V., Rahman Siddiquee, M.M., Feng, R., Tajbakhsh, N., Gotway, M.B., Liang, J.: Models genesis: Generic autodidactic models for 3D medical image analysis. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 384–393. Springer International Publishing, Cham (2019)
29. Zhuang, X., Li, Y., Hu, Y., Ma, K., Yang, Y., Zheng, Y.: Self-supervised feature learning for 3D medical images by playing a Rubik's cube. In: MICCAI (2019)

A Supplementary Material

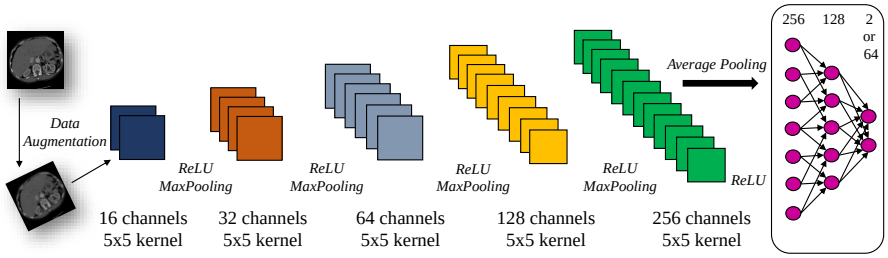


Fig. 3: The proposed TinyNet used in our experiments.

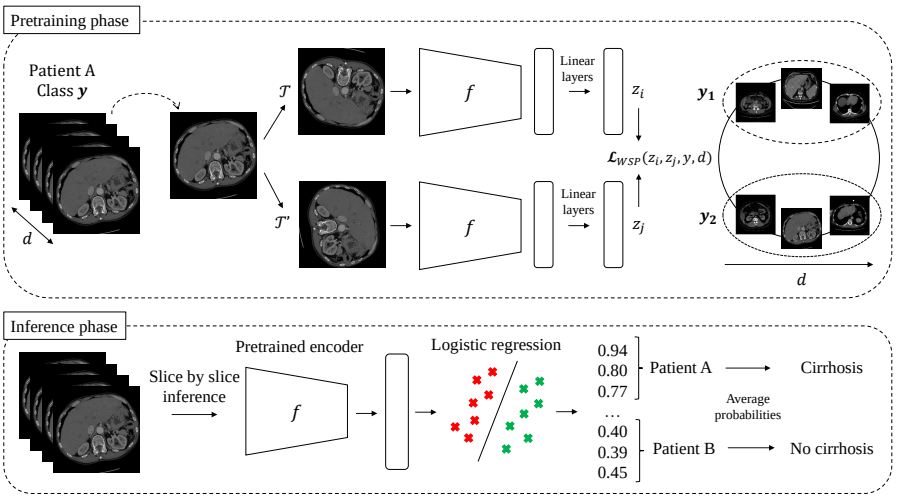


Fig. 4: The full workflow of our method.

Table 3: Resulting 5-fold cross-validation AUCs of the proposed method using the TinyNet backbone, varying the value of σ. In the paper, we chose the value of $\sigma = 0.1$. One can interpret σ as the proportion of slices around the anchor with a high weight accordance. The higher the value of σ is, the more slices will be assigned a high weight value.

$\boldsymbol{\sigma}$	$\boldsymbol{\mathcal{D}^1_{histo}}$ (N=106)	$\boldsymbol{\mathcal{D}^2_{histo}}$ (N=49)	$\boldsymbol{\mathcal{D}^{1+2}_{histo}}$ (N=155)
0.01	0.81 (±0.07)	0.85 (±0.13)	**0.81 (±0.08)**
0.1	**0.85 (±0.10)**	**0.91 (±0.11)**	0.79 (±0.10)
0.2	0.75 (±0.08)	0.85 (±0.07)	0.72 (±0.08)
0.3	0.78 (±0.09)	0.82 (±0.31)	0.76 (±0.06)
0.5	0.73 (±0.15)	0.89 (±0.12)	0.76 (±0.07)

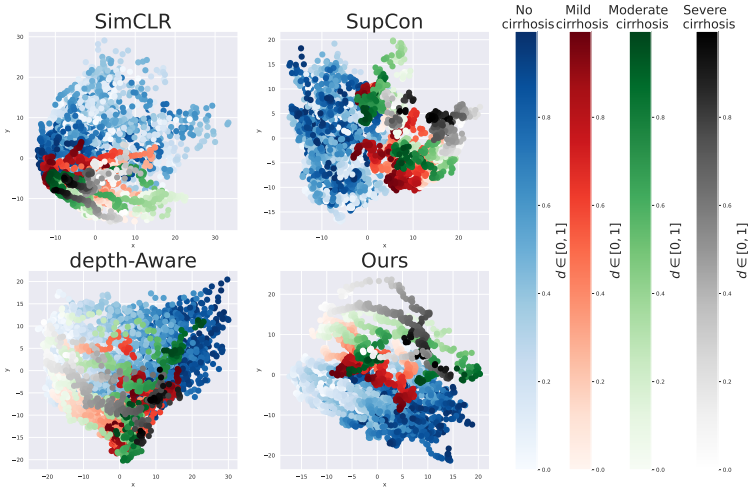
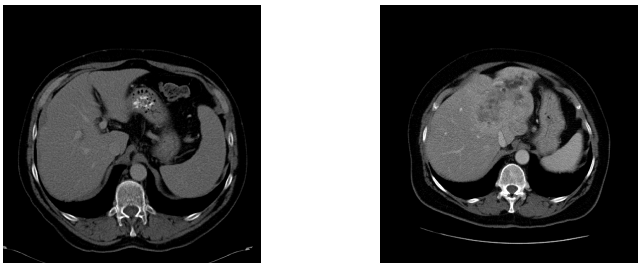


Fig. 5: Projections of the ResNet-18 representation vectors of 70 randomly selected subjects of the training set of $\mathcal{D}_{radio}$ onto the first two modes of a PCA. Each dot represents a 2D slice. Color gradient refers to different depth positions. SimCLR and SupCon provide a remarkable separation between the healthy subjects (in blue) and the rest. However, classes mild moderate and severe are hardly separated. depth-Aware reaches an interesting global color gradient, but struggles to separate the cirrhotic classes. Our method provides the best class separation and at the same time correctly encodes the depth position.



(a) Cirrhotic case with the highest probability predicted by WSP.

(b) False negative misclassified by all the methods.

Fig. 6: CT slices from $\mathcal{D}^2_{histo}$. On the left, the proposed method predicts the highest probability with 0.53 while SupCon, depth-Aware and SimCLR predict 0.51, 0.50 and 0.47 respectively. On the right, a false negative case predicted by all the models, possibly due to the slightly smaller size of the slice.