Deep learning methods for segmentation of images of frozen tissue sections

Dario Sitnik, mag. ing.
(dario.sitnik@irb.hr)
mentor: nasl. izv. prof. dr. sc. Ivica Kopriva

Faculty of Electrical Engineering and Computing, University of Zagreb

Introduction

According to the International Agency for Research on Cancer of the World Health Organization, in 2020, colorectal cancer was the third most common cancer type worldwide, causing:

- 2 million new cases,
- and 1 million cancer-related deaths per year.

The gold standard for diagnosing cancer is a microscopic analysis of stained histopathological specimens made by pathologists. As cancer incidence rises, pathologists' tasks become more complex and time-consuming. That's why computer-aided diagnosis (CAD) is supposed to alleviate pathologists' workload.

Problem description

Staining of histopathological specimens improves tissue structure visibility. Hematoxylin and eosin (H&E) are the most often used dyes in pathology. Formalin-fixed paraffin embedding (FFPE) fixes a tissue sample on the glass slide. Preparing a glass slide for microscopy takes around 48 hours. In time-critical applications, such as intraoperative diagnosis, FFPE is unacceptable. Instead, a tissue fast freezing, followed by cutting on a cryotome, and staining with H&E is employed (see Fig. 1). However, the quality of the stained image is affected by various batch effects. Thus, having a robust CAD-aided intraoperative decision-making system that can serve as a *second reader* is desirable.

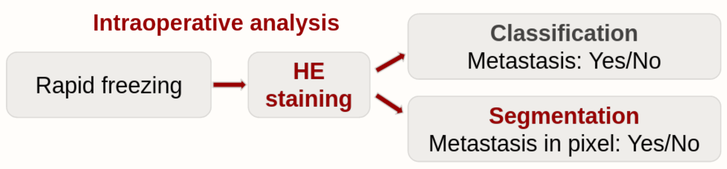

Figure 1: Intraoperative diagnosis procedures

However, only a few studies have used frozen sections. The major problem is the lack of publicly available datasets with enough annotated histopathology pictures to train classification systems. This is especially true for pixel-level annotations.

Materials and Methods

Staining and image acquisition

The pathologist diagnosed metastatic colon cancer in the liver using primary antibodies specific to hepatocyte protein, a transcription factor expressed in colorectal carcinoma cells, and cytokeratin 20 as an adenocarcinoma cell marker. Images were collected at a 400× magnification. An H&E sample image can be seen in Fig. 2.

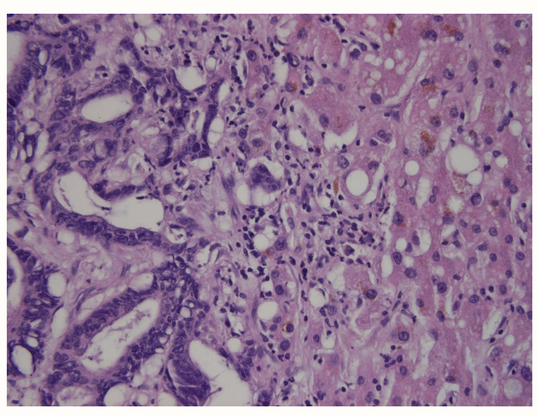

Figure 2: Image of H&E stained section of adenocarcinoma of a colon in a liver.

Pixel-wise labeling

Four pathologists, two residents in pathology and one final-year medical student labeled images of H&E stained section images pixel-by-pixel. For the labeling, a super-pixels based approach was used to group similar pixels in a magnified region. Annotators might pick between the super-pixel and brush tools. Different super-pixel algorithms, e.g. SLIC, Watershed, Quickshift, and Felzenszwalb from the *scikit-image* package, could be chosen.

Pathologists might refine or reject indicated areas throughout the annotation process. While the software system reduced annotators' workload by generating super-pixels, they could still zoom in and annotate the image at the pixel level (e.g. Fig. 3). Seven experts were used for pixels annotation because of inter-observer variability and subjectivity. To measure inter-annotator agreement, we used Fleiss' kappa statistics.

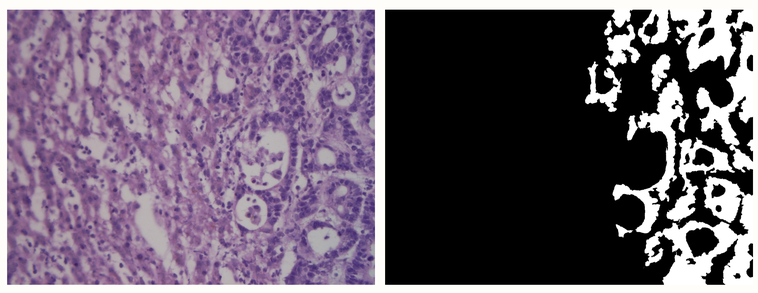

Figure 3: H&E stained section and corresponding ground truth obtained by the majority vote.

Stain normalization

On pathologist-selected target images, we employed structure-preserved color (stain) normalization [1] to adjust for color variations caused by experimental variations due to the slide preparation process. Thus, machine learning models may be trained on either the H&E stained frozen sections or on one of the two stain-normalized versions of datasets (see Fig. 4). This work, however, integrates diagnostic results from all three datasets by majority voting.

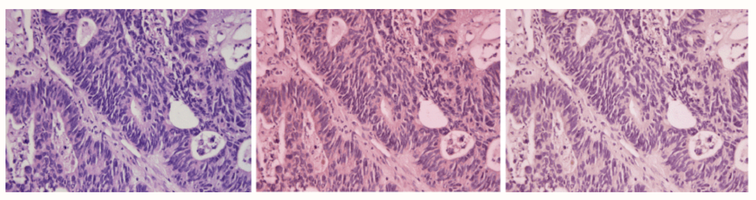

Figure 4: Images of H&E stained section (left) and stain-normalized versions with respect to target 1 (middle) and target 2 (right).

Conventional machine learning and deep learning models

SVM and kNN are representative examples of machine learning classifiers that can perform well in diagnostic tasks. The *incrementalLearner* function from Matlab 2020b was used to perform linear SVM's incremental learning. The sliding window algorithm (ADWIN) is used to apply kNN adaptively. We also applied DeepLabv3+, UNet, and UNet++ as deep learning models for segmentation. They were implemented in Python using the Pytorch framework.

Results

In Table 1, performance scores of all models were presented. It can be seen that there is a major improvement when comparing deep learning vs. conventional machine learning models. However, the difference between deep models is less significant whether their weights are trained from scratch or pretrained on a classification problem of natural images.

Table 1: Performance scores of five different models on CoCaHis datasets. Each model is trained from the scratch (random initialization) and also finetuned after initialization on ImageNet classification problem. Best results are in bold.

	Metric	Raw		SN (target 1)		SN (target 2)		Majority Vote	
		from scratch	pretrained	from scratch	pretrained	from scratch	pretrained	from scratch	pretrained
SVM	BACC	0.6869	-	0.7640	-	0.7516	-	0.7594	-
	F1	0.4235	-	0.6830	-	0.7035	-	0.6698	-
	PREC	0.3931	-	0.5279	-	0.5467	-	0.5234	-
kNN	BACC	0.6971	-	0.7657	-	0.6885	-	0.7295	-
	F1	0.5597	-	0.6434	-	0.5513	-	0.5935	-
	PREC	0.4355	-	0.5523	-	0.3990	-	0.4636	-
DeepLabv3+	BACC	0.8610	0.8852	0.8675	0.8639	0.8628	0.8491	0.8775	0.8825
	F1	0.7884	**0.8114**	0.8107	0.7978	0.8011	0.7822	0.8219	0.8225
	PREC	**0.7559**	0.7510	**0.8183**	0.7809	**0.7993**	0.7833	**0.8193**	0.8006
UNet	BACC	0.8630	0.8617	0.8598	0.8673	0.8410	0.8634	0.8746	0.8886
	F1	0.7855	0.7780	0.8161	0.8045	0.7737	0.8013	0.8155	0.8316
	PREC	0.7371	0.7212	0.8103	0.7934	0.7867	0.7976	0.8063	0.8108
UNet++	BACC	0.8851	**0.8862**	**0.8792**	0.8632	0.8597	**0.8656**	0.8841	**0.8934**
	F1	0.7874	0.8097	**0.8211**	0.8018	0.7961	**0.8017**	0.8274	**0.8367**
	PREC	0.7500	0.7451	0.8091	0.8001	0.7928	0.7893	0.8131	0.8111

Conclusion

It is expected that artificial intelligence and computational pathology can assist pathologists in detection and grading of cancer. A number of obstacles stand in the way of the development of CAD-assisted intraoperative decision-making systems. Due to variations in image quality and the time needed, the pixel-level annotations vary greatly. Thus, multiple annotations are required to get a reasonable estimate of the most likely outcome of the annotation procedure. To that end, we created a pixel-level annotated database containing 82 histopathological images of H&E stained frozen sections from 19 patients with adenocarcinoma of a colon in a liver. The ground truth maps were generated using majority voting among experts.

On an independent test set, deep learning classifiers outperformed SVM and kNN classifiers in terms of balanced accuracy, F_1 score, and precision in the amounts of 14%, 15% and 26%, respectively. Both U-Net and U-Net++ classifiers performed within 2% of each other, whether trained from scratch or pre-trained on the image dataset of different domain [2].

Project Acknowledgement

This work has been supported in part by the Croatian Science Foundation Grant IP-2016-06-5235, and in part by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

References

[1] A. Vahadane et al., "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," in IEEE Transactions on Medical Imaging, vol. 35, no. 8, p. 1962-1971, 2016

[2] Sitnik, Dario, et al. "A dataset and a methodology for intraoperative computer-aided diagnosis of a metastatic colon cancer in a liver." Biomedical Signal Processing and Control vol. 66, p. 102402, 2021