

به نام خدا

گزارش کار کارآموزی تابستان ۱۴۰۱

شرکت زیستارن

پارسا حسینی

sp.hoseiny@gmail.com

چکیده

دوره تابستانی کارآموزی با هدف تحقیق و بررسی هدف اصلی دوره بررسی و تحلیل عکس‌های مربوط به سلول و بافت خونی بوده است. در ادامه کدها و مدل‌هایی که روی دیتای سلول خونی شرکت اجرا شدند یا برای اجرا مناسب هستند آورده شده است.

شرح کار

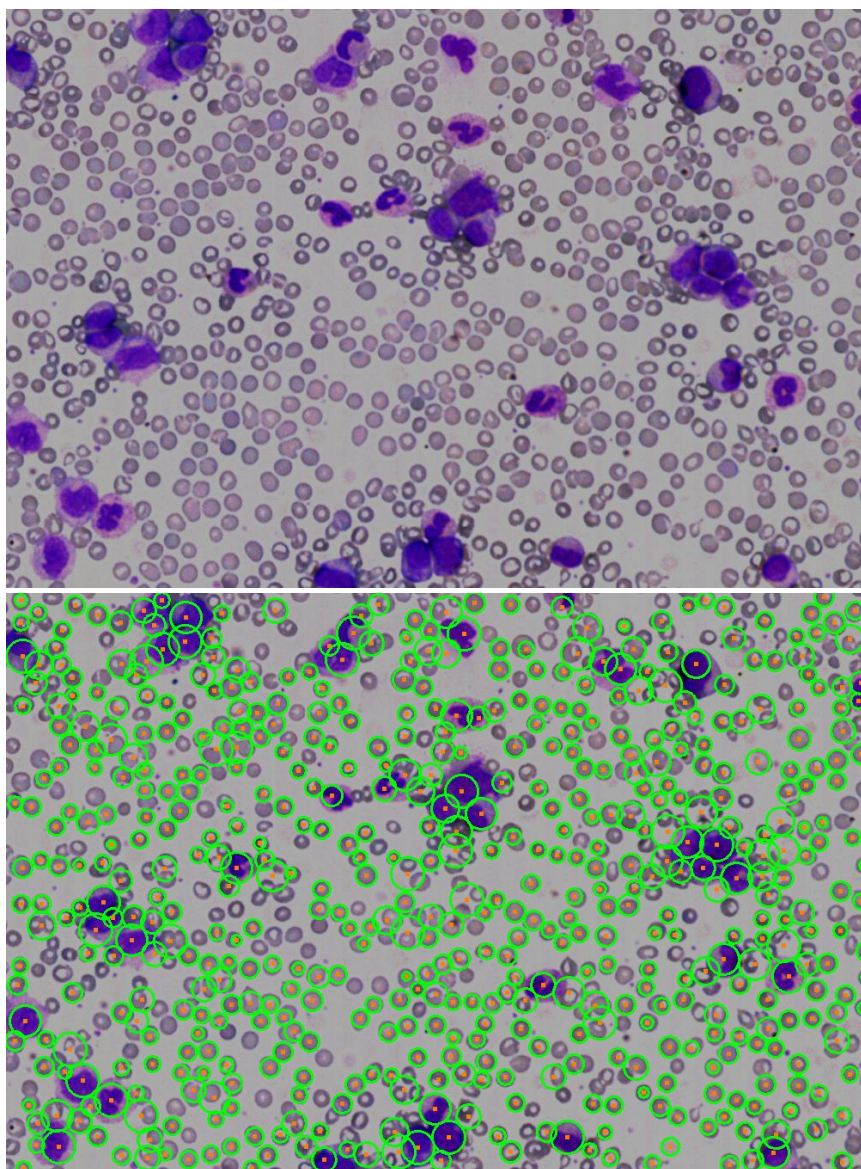
ابتدا مدتی به آشنایی به مسائل حوزه دیجیتال، تسک‌های مختلف و دیتاست‌های موجود سپری شد. در میان منابع متنوع، یک survey (لینک ۹) ارزش بررسی دوباره در آینده را دارد. این survey حوزه پاتولوژی دیجیتال به خوبی توضیح داده شده است.

Table 5.1: Public Datasets for Image Analysis Tasks in Digital Pathology.

Dataset	Task	#Images
Arganda et al. [13]	Neuron boundary segmentation	60
ISBI2012-EM [4]	Neuron segmentation	30
GlaS [82]	Colon gland segmentation	160
AMIDA2013 [93]	Mitosis detection	
ICPR 2012 [27]	Mitosis detection	50
ICPR 2014 [27]	Mitosis detection	50
TUPAC [9]	Tumor detection	300
Camelyon16 [18]	Metastasis detection	299
Camelyon17 [18]	Metastasis detection	400
TMA Thyroid [8]	Outcome prediction	
BACH [6]	Tissue subtypes classification	400
TCIA [2]	Multiple	—
TCGA [1]	Multiple	—
Her2 [71]	Her-2 Scoring	172
Cellavision [3]	Cell segmentation	100
Enjoypath [5]	Multiple	318
PCam [7]	Metastasis detection	327680

۱- دیتاست‌های پابلیک موجود در زمینه پاتولوژی دیجیتال به همراه اندازه‌شان (برگرفته از لینک ۹)

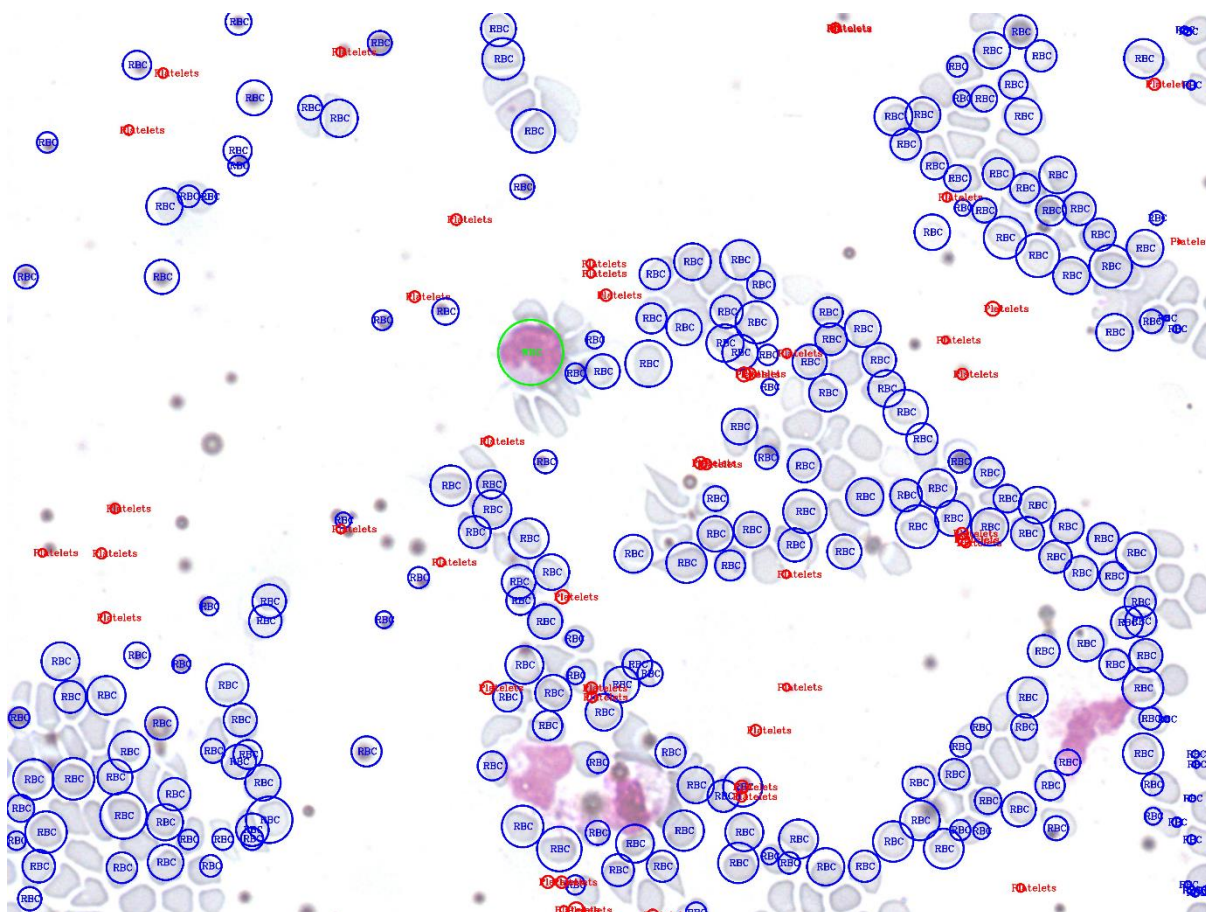
همچنین با کمک کتابخانه‌های آماده پردازش تصویر پایتون (بدون یادگیری) هم برنامه اولیه‌ای روی یکی از نمونه‌های شرکت اجرا شد. یک نمونه را می‌توانید در عکس زیر ببینید. کد این قسمت روی گیت نهایی (لینک ۱) قرار داده شده است.



۲- نمونه خروجی کد ابتدایی پردازش تصویر پایتون. (لینک ۶)

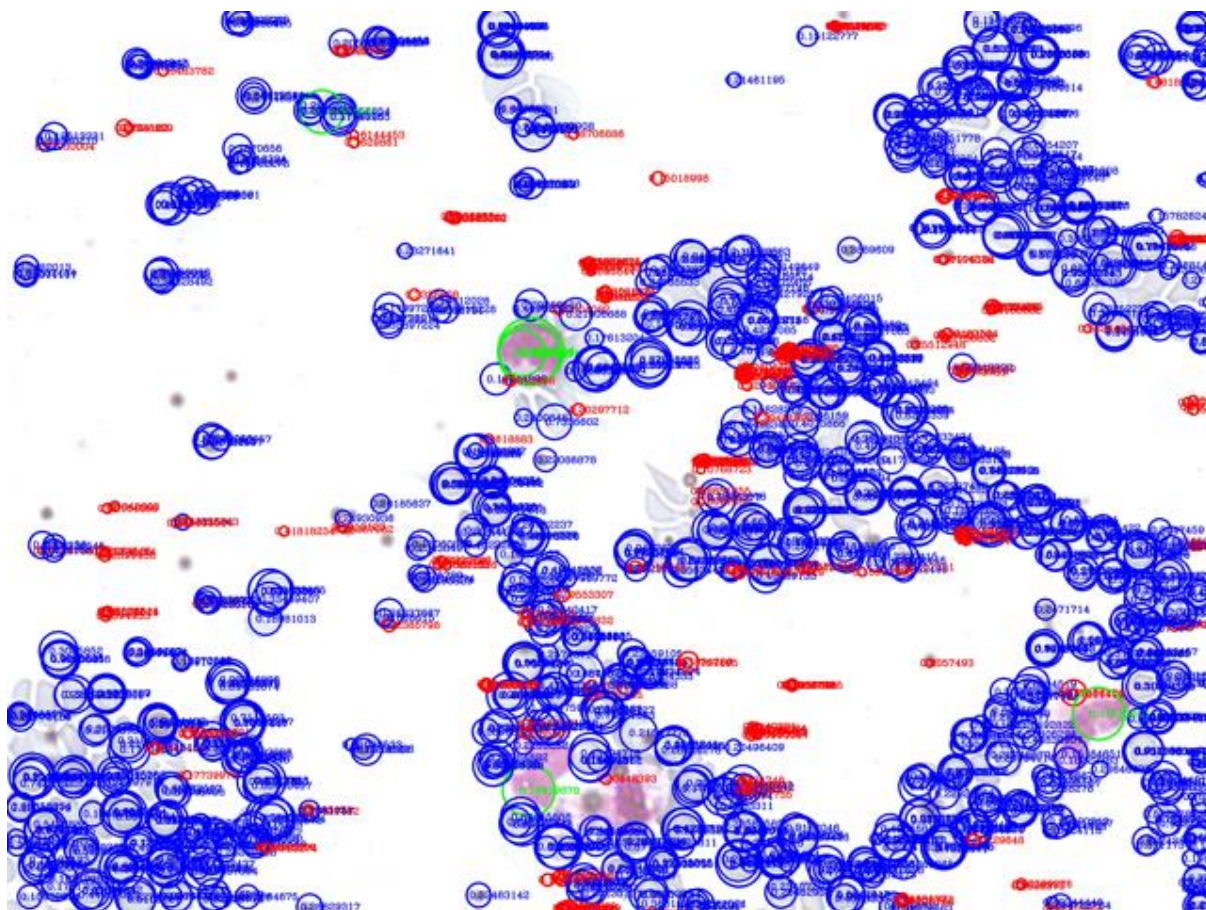
مسئله دیگری که در دیتای شرکت وجود دارد، حجم بالای عکس‌ها است. یک سمپل شرکت ممکن است حتی چند گیگ باشد و برای همین نمونه‌گیری از این عکس‌ها خود مسئله است. در لینک ۱ یک نوتبوک ساده برای اینکار قرار داده شده است. البته ما از کد دیگری هم که قبلاً در شرکت زده شده بود برای اینکار استفاده کردیم که به دلیل محرمانگی جایی عمومی قرار ندادیم.

لینک ۲ یکی از مدل‌های مهمی است که توانستیم روی دیتای خودمان اجرا کنیم. کد تغییر داده شده و نوتبوک مورد نظر هر دو روی گیت نهایی وجود دارند. برای اینکه عکس‌هایمان را بهتر تشخیص دهد، باید رنگ‌های آن را شبیه به رنگ‌های دیتاستی که مدل روی آن آموزش دیده است می‌کردیم. برای اینکار RGB هر پیکسل را با یک تبدیل خطی ساده طوری کردیم که پس‌زمینه سفید شود. اینکار تاثیر بسیار زیادی در کیفیت خروجی‌هایمان داشت.



۳- نمونه خروجی مدل روی عکس ورودی. دقت کنید که مدل هم وجود سلول را تشخیص می‌دهد، شعاع تقریبی آن را می‌گوید و نوع آن را بین ۳ نوع گلبول سفید، گلبول قرمز و پلاکت مشخص می‌کند.

با آزمایش مشخص شد که خروجی مدل بسیار به تغییرات رنگی ما و اندازه زوم عکس ورودی وابسته است و خروجی آن با کمی تغییر در این پارامترها خیلی عوض می‌شود. برای همین ایده‌ای که داشتیم این بود که عکس با پارامترهای مختلف رنگی و زوم ورودی داده شده، و به نوعی خروجی‌ها با هم ترکیب شوند که خروجی نهایی موارد اشتباه یک حالت را اصلاح کند. یک کد اولیه از این ایده روی گیت وجود دارد. برای مثال در عکس زیر نمونه بالا با زوم‌های مختلف به مدل ورودی داده شده و خروجی همه با هم کاملاً ترکیب شده است. در صورتی که یک صحت سنجی ساده روی این عکس انجام شود می‌توان دید که تقریباً همه سلول‌ها را تشخیص می‌دهد.



۴- یک نمونه از ترکیب کردن ساده. می‌توان با قرار دادن یک ترشهولد موارد یکسان را حذف کرد. کد این مورد هم روی گیت موجود است.

در انتها با توجه به اینکه هدف نهایی لیبل زدن تعدادی از عکس‌های شرکت و بعد transfer learning است، دیدن لینک ۸ خالی از لطف نیست. خیلی از پروژه‌های این حوزه از این ابزار برای لیبل زدن عکس‌هایشان استفاده کرده‌اند.

لینک‌های مهم

1. <https://github.com/sudoparsa/Zistagene>
2. <https://github.com/MahmudulAlam/Automatic-Identification-and-Counting-of-Blood-Cells>
3. <https://github.com/MahmudulAlam/Complete-Blood-Cell-Count-Dataset>
4. https://github.com/abhanjac/blood_cell_image_segmentation
5. <https://pathml.readthedocs.io/en/latest/overview.html>
6. <https://www.geeksforgeeks.org/python-blood-cell-identification-using-image-processing/>
7. <https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/>

8. <https://github.com/tzutalin/labelImg>
9. <https://arxiv.org/abs/1910.12329>