**تشخیص اعداد دست نویس اسلب**

**الهام هاتفی**

به منظور تشخیص اعداد دست نویس اسلب دو تکنیک مورد استفاده قرار گرفت. تکنیک اول ترکیب پردازش تصویر و یادگیری عمیق بود که برای شرایط مختلف نورپردازی مختلف الگوریتم را با چالش مواجه کرد. تکنیک دوم مبتنی بر شبکه‌های عصبی عمیق (شبکه‌ی یولو) بود به صورتی که علاوه بر دسته‌بندی داده‌ها، موقعیت هر یک از داده‌ها را نیز مشخص می‌کرد. مزیت این روش نسبت به روش قبلی در این بود که علاوه بر آن که به تغییرات نور‌پردازی مقاوم شد، مکان اعداد در تصویر و دسته‌بندی آنها را تنها با استفاده از شبکه‌ی عصبی مذکور و در یک مرحله انجام داد. در ادامه جزئیات این دو روش تشریح می‌گردد:

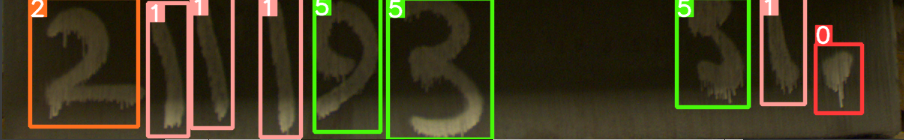
**تکنیک اول:** همان طور که بیان شد تکنیک اول ترکیب پردازش تصویر کلاسیک و یادگیری عمیق بود به صورتی که ابتدا یک آستانه‌گیری برای تصاویر در نظر گرفته شد، تا رقم‌های تصاویر مشخص گردند و سپس از تکنیک پیدا کردن کانتور(find-contours) استفاده شد تا رقم های تصاویر استخراج شوند. سپس رقم‌های استخراج شده برای هر تصویر به یک شبکه‌ی عصبی آموزش دیده شده ارسال شد تا برای هر رقم یک دسته‌بندی انجام شود. به منظور پیش‌بینی ارقام از یک شبکه‌ی عصبی کانولوشنی استفاده شد به صورتی که این شبکه‌ی عصبی بر مبنای اعداد دست‌نویس mnist آموزش دیده شده بود.

**معایب روش** : مشکلی که این روش داشت، در شرایط مختلف نورپردازی نیاز به گرفتن ترشلدهای مختلف برای تصاویر بود. هم چنین با توجه به این که عدد صفر موجود در داده های mnist متفاوت از داده‌های اعدادد دست نویس بود دقت بالائی برای عدد صفر حاصل نشد.

**تکنیک دوم :** همان‌طور که بیان شد در تکنیک دوم، از شبکه‌ی عصبی یولو استفاده شد. به منظور آموزش این شبکه‌ی عصبی، تصاویر مرتبط با اعداد دست نویس اسلب که از فولاد مبارکه گردآوری شده بود مورد استفاده قرار گرفت. برای استفاده از این تصاویر می‌بایست، موقعیت اعداد در تصویر به همراه برچسب اعداد مشخص می‌گردید. در واقع برای آموزش شبکه‌ی یولو تصاویر می بایست برچسب‌گذاری می‌شدند. به منظور برچسب گذاری تصاویر ازlabel image استفاده شد، که ابتدا با فرمت XML برچسب گذاری انجام شد، و بعد ا ز آن فرمت XML به TXT تبدیل شد. در این تکنیک از دو رویکرد استفاده گردید. رویکرد اول به منظور تشخیص اعداد برای هر اسلب بود و رویکرد دوم برای تشخیص اعداد برای تصاویر کامل بود. در ادامه این دو رویکرد تشریح می گردد:

**رویکرد اول:**

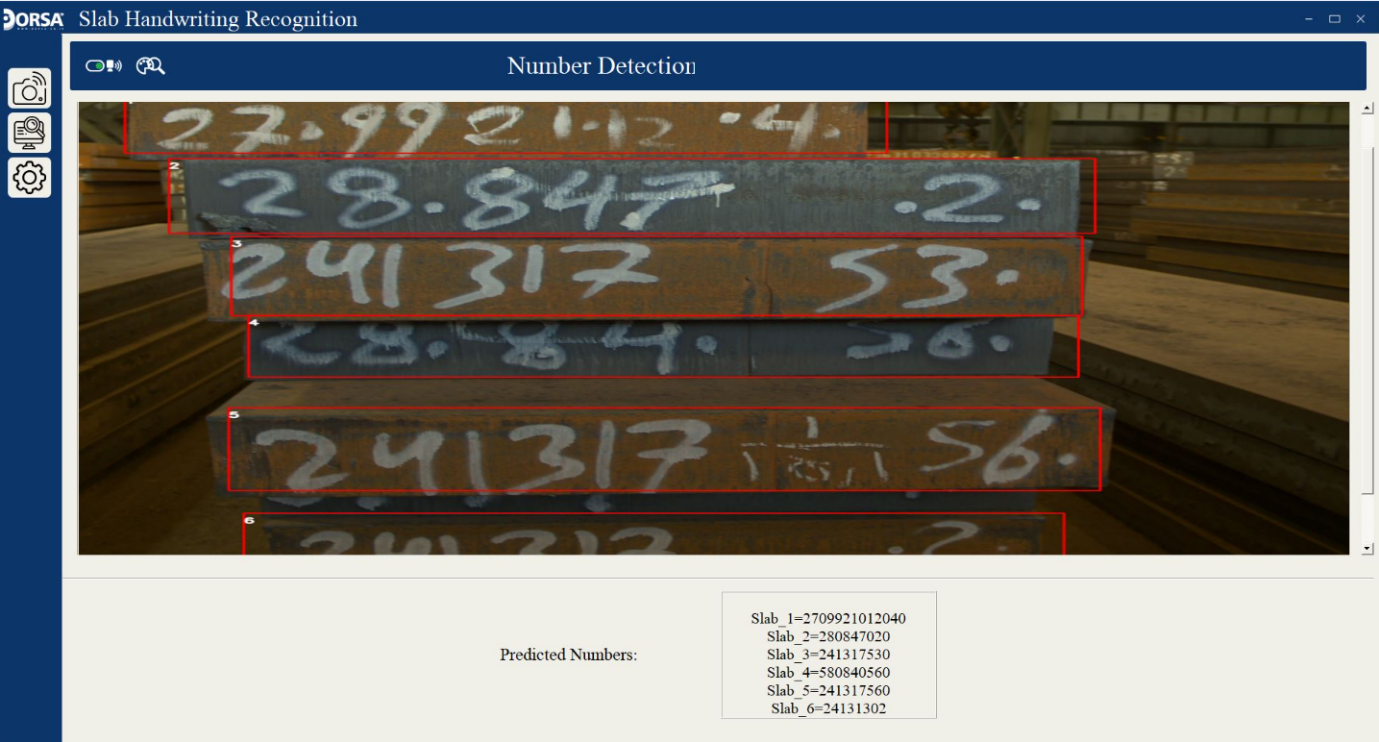
در این رویکرد ابتدا تصویر هر اسلب، از تصاویر کامل جدا شد و برچسب‌گذاری انجام شد به منظور تشخیص موقعیت اعداد و برچسب متناظر هر عدد، از یک شبکه‌ی یولو استفاده شد. برای آموزش شبکه، از تعداد epoch 400، با سایز batch 2 ، استفاده گردید به صورتی که نسبت تصاویر train به تصاویر validation ، 80 به 20 بود. در ادامه تشخیص اعداد برای این تصاویر نشان داده شده است.

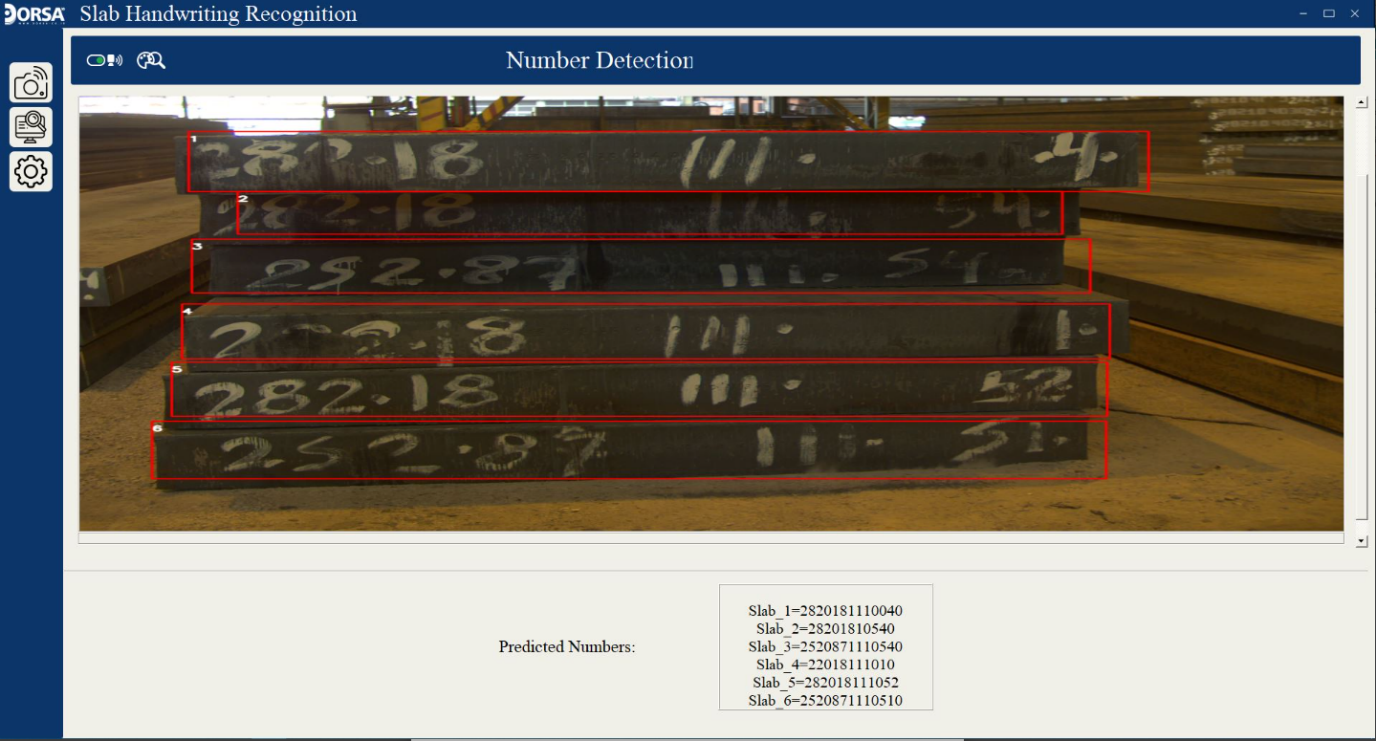
****

****

**رویکرد دوم:**

در این رویکرد تشخیص اعدا دست‌نویس اسلب برای کل تصویر صورت پذیرفت. به صورتی که برای انجام این کار از دو شبکه‌ی عصبی یولو استفاده شد. هدف استفاده از شبکه‌ی عصبی اول، پیدا کردن موقعیت هر اسلب در تصویر بود. بدین منظور برچسب گذاری برای تصاویر کامل به صورتی انجام شد که موقعیت هر اسلب مشخص شد و برچسب صفر برای آن در نظر گرفته شد. برای آموزش شبکه، از تعداد epoch 400 ، با سایز batch ، 2 استفاده گردید به صورتی که نسبت تصاویر train به تصاویر validation ، 80 به 20 بود. بعد از مشخص شدن موقعیت اسلب در هر تصویر، این تصاویر به شبکه‌ی یولو آموزش دیده شده در رویکرد اول داده شد و برچسب و موقعیت اعداد برای هر اسلب حاصل شد. طراحی ui نیز برای دو رویکرد اول و دوم صورت پذیرفت که در ادامه برای تصاویر کامل نشان داده شده است:





آموزش شبکه :

**برای تصاویر تکی**

**epoch 100 , batch 2 detect21- 400 epoch**

**detect 23(adding 3 and 6) epoch 100 , batch 2**

**برای تصاویر کلی**

**epoch 300 , batch 2**