



دانشگاه اصفهان

دانشکده مهندسی‌کامپیوتر

**پایان‌نامه کارشناسی**

رشته مهندسی کامپیوتر

بسته اصلی: هوش مصنوعی و رباتیکز

بسته فرعی: نرم افزار

**عنوان**

**پویشگر و تحلیلگر ارزش مواد غذایی**

**استاد راهنما:**

سرکارخانم دکتر افسانه فاطمی

**پژوهشگر:**

علی پورقیصری

شهریور 1403



دانشگاه اصفهان

دانشکده مهندسی‌کامپیوتر

پروژه کارشناسي رشته‌ي مهندسي کامپیوتر

**آقاي علی پورقیصری**

**تحت عنوان**

**پویشگر و تحلیلگر ارزش مواد غذایی**

در تاريخ / / 14 توسط هيأت داوران زير بررسي و با نمره به تصويب نهايي رسيد.

1. استاد راهنماي پروژه

دکتر امضا

1. استاد داور

دکتر امضا

امضای‌ مدیرگروه

**تشکر و قدرداني**

سپاس بیکران از پروردگار که بار دیگر فرصت یادگیری را به من ارزانی داشت. از پدر و مادر عزیز و مهربانم که در طول زندگی با تحمل زحمات فراوان، همواره برای موفقیت و خوشبختی من تلاش کردند، صمیمانه قدردانی می‌کنم. همچنین از استاد ارجمند، دکتر افسانه فاطمی، که با تلاش‌های بی‌دریغ خود در تحقق این پایان‌نامه سهم بسزایی داشتند، نهایت تشکر را دارم. قدردان دوستان عزیزم نیز هستم که با حمایت‌هایشان در به سرانجام رساندن این پایان‌نامه نقش مهمی ایفا کردند.

لازم می‌دانم از تمامی افرادی که در بهبود و توسعه هوش مصنوعی در زبان فارسی سهمی داشته‌اند، تشکر کنم و امیدوارم که با تلاش توسعه‌دهندگان، دیگر هیچ محدودیتی برای انجام پژوهش‌های فارسی در مقایسه با سایر زبان‌ها وجود نداشته باشد.

**تقدیم به**

**پدر و مادر عزیزم**

**چکیده:**

در این پروژه، یک برنامه کاربردی توسعه داده شده است که به کاربران امکان می‌دهد با استفاده از تکنیک‌های پیشرفته پردازش تصویر و بینایی ماشین، اطلاعات مهمی را از روی برچسب‌های ارزش غذایی محصولات فروشگاهی استخراج و خوانده شوند. این برنامه با پویش برچسب‌های موجود بر روی محصولات، اطلاعاتی نظیر میزان انرژی محصول برحسب کیلوکالری، و مقدار مواد مختلف از جمله قند، نمک، چربی، و اسیدهای چرب ترانس برحسب گرم را نمایش می‌دهد. علاوه بر این، برنامه توانایی تشخیص و ارائه مقدار ماده معیار در محصول را برحسب گرم یا میلی‌لیتر دارد، که می‌تواند به کاربران در تحلیل بهتر ارزش غذایی محصول کمک کند.

در مرحله نهایی، برنامه شماره پروانه بهداشتی ساخت محصول را نیز تشخیص می‌دهد و با استعلام این شماره، کاربران قادر خواهند بود به اطلاعات فنی دقیق‌تری همچون نام تجاری محصول، نام کارخانه تولیدکننده، تاریخ دریافت و انقضای پروانه بهداشتی، و آدرس کارخانه دسترسی پیدا کنند. این قابلیت به ویژه برای مصرف‌کنندگانی که به دنبال اطلاعات دقیق و موثق درباره محصولات مصرفی خود هستند، بسیار کارآمد و مفید خواهد بود. هدف اصلی این پروژه، افزایش آگاهی مصرف‌کنندگان در انتخاب محصولات سالم‌تر و بهبود کیفیت خرید آنها با ارائه اطلاعات شفاف و قابل اعتماد می‌باشد.

فهرست مطالب

|  |  |
| --- | --- |
| صفحه | عنوان |

[فصل اول 8](#_Toc174961477)

[مقدمه و بیان مسئله 8](#_Toc174961478)

[۱-۱- توضیحات کلی 8](#_Toc174961479)

[۲-۱- اهداف پروژه 9](#_Toc174961480)

[1-2-1- افزایش آگاهی مصرف‌کنندگان از محتوای تغذیه‌ای محصولات غذایی 9](#_Toc174961481)

[1-2-2- تسهیل فرآیند انتخاب محصولات غذایی سالم و مناسب 9](#_Toc174961482)

[1-2-3- ارائه اطلاعات فنی و بهداشتی محصولات 10](#_Toc174961483)

[1-2-4- افزایش شفافیت در صنعت غذا 10](#_Toc174961484)

[1-2-5- حمایت از تولیدکنندگان متعهد به کیفیت 10](#_Toc174961485)

[1-2-6- بهبود سلامت عمومی جامعه 10](#_Toc174961486)

[۳-۱- دامنه و محدوده پروژه 11](#_Toc174961487)

[1-3-1- جمع‌آوری و آماده‌سازی داده‌ها 11](#_Toc174961488)

[1-3-2- عملکرد برنامه در محیط واقعی 11](#_Toc174961489)

[1-3-3- دامنه پروژه 12](#_Toc174961490)

[1-3-4- خارج از دامنه پروژه 13](#_Toc174961491)

[1-4- چالش ها 14](#_Toc174961492)

[فصل دوم 16](#_Toc174961493)

[ادبیات پژوهش 16](#_Toc174961494)

[۱-۲- مقدمه: 16](#_Toc174961495)

[فصل سوم 17](#_Toc174961496)

[پیشینه پژوهش 17](#_Toc174961497)

[۱-3- مقدمه: 18](#_Toc174961498)

[۲-3- تاریخچه و توسعه OCR: 18](#_Toc174961499)

[3-3-مروری بر مطالعات مرتبط: 19](#_Toc174961500)

[4-3- جمع بندی: 21](#_Toc174961501)

[ ۱-۲-۲- تراز پلاک: 21](#_Toc174961502)

[ ۲-۲-۲- برنامه برچسب‌گذاری پلاک: 22](#_Toc174961503)

[ ۳-۲-۲- رفع نویز و شفاف‌سازی پلاک: 22](#_Toc174961504)

[ ۴-۲-۲- دنبال‌کننده پلاک: 22](#_Toc174961505)

[ ۵-۲-۲- انتقال برنامه‌های نوشته‌شده به محیط‌کاری دیپ‌استریم: 22](#_Toc174961506)

[۱-۳- مقدمه: 24](#_Toc174961507)

[۲-۳- تجارب: 24](#_Toc174961508)

[۳-۳- دست‌آوردهای فنی: 24](#_Toc174961509)

[۴-۳- موفقیت‌ها و ناکامی‌ها: 25](#_Toc174961510)

[۵-۳- ارزیابی: 26](#_Toc174961511)

[فصل چهارم 27](#_Toc174961512)

[گزارش‌های فنی 27](#_Toc174961513)

[۱-۴- مقدمه: 27](#_Toc174961514)

[۲-۴- تراز‌سازی پلاک: 27](#_Toc174961515)

[۱-۲-۴- کشف کناره‌ها: 28](#_Toc174961516)

[۲-۲-۴- الگوریتم پیدانمودن مستطیل: 28](#_Toc174961517)

[۳-۲-۴- تراز‌سازی توسط شبکه‌های عصبی عمیق: 29](#_Toc174961518)

[۴-۲-۴- شاخص آردو: 31](#_Toc174961519)

# فصل اول

# مقدمه و بیان مسئله

## ۱-۱- توضیحات کلی

در دنیای امروز که سرعت زندگی و مشغله‌های روزمره افزایش یافته، اهمیت دسترسی به اطلاعات دقیق و قابل اطمینان درباره محصولات غذایی بیشتر از همیشه احساس می‌شود. مصرف‌کنندگان به دنبال اطلاعاتی هستند که به آنها کمک کند تا با آگاهی کامل‌تر محصولات غذایی مناسب‌تر و سالم‌تری را انتخاب کنند. این موضوع به ویژه با توجه به افزایش نگرانی‌ها درباره تأثیرات مضر مواد غذایی نظیر قند، نمک، چربی‌های اشباع و اسیدهای چرب ترانس بر سلامت انسان، اهمیت بیشتری پیدا کرده است.

یکی از چالش‌های اساسی مصرف‌کنندگان در این زمینه، عدم دسترسی سریع و آسان به اطلاعات دقیق و جامع درباره ارزش غذایی محصولات است. برچسب‌های مواد غذایی که به طور معمول بر روی محصولات درج می‌شوند، به دلیل پیچیدگی اطلاعات یا اندازه کوچک نوشته‌ها، ممکن است برای بسیاری از مصرف‌کنندگان خوانا نباشند یا به راحتی قابل فهم نباشند. علاوه بر این، عدم آگاهی کافی از محتوای دقیق محصولات می‌تواند منجر به انتخاب‌های نادرست و در نهایت تهدید سلامت مصرف‌کنندگان شود.

از سوی دیگر، نیاز به دسترسی به اطلاعات فنی محصولات، نظیر نام تجاری، کارخانه تولیدکننده، و تاریخ انقضا، از اهمیت بسزایی برخوردار است. این اطلاعات نه تنها به مصرف‌کنندگان کمک می‌کند تا از کیفیت و ایمنی محصولات مطمئن شوند، بلکه امکان پیگیری و استعلام شماره پروانه بهداشتی ساخت محصول را نیز فراهم می‌آورد.

پروژه حاضر با هدف ارائه راه‌حلی نوآورانه برای این چالش‌ها طراحی شده است. در این پروژه، یک برنامه کاربردی توسعه داده شده است که به کاربران امکان می‌دهد با پویش برچسب‌های ارزش غذایی محصولات فروشگاهی، اطلاعات مندرج بر روی آنها را با استفاده از تکنولوژی پردازش تصویر[[1]](#footnote-1) و بینایی ماشین[[2]](#footnote-2) به صورت خودکار بخوانند. این برنامه قادر است میزان انرژی، قند، نمک، چربی‌ها، اسیدهای چرب ترانس و سایر اطلاعات مهم تغذیه‌ای را استخراج کرده و به شکلی ساده و قابل فهم در اختیار کاربران قرار دهد. علاوه بر این، امکان استعلام شماره پروانه بهداشتی ساخت محصول نیز فراهم شده است که به کاربران اجازه می‌دهد اطلاعات فنی محصول را بررسی کرده و از صحت و سلامت آن اطمینان حاصل کنند.

این پروژه با هدف تسهیل دسترسی مصرف‌کنندگان به اطلاعات تغذیه‌ای و فنی محصولات غذایی، افزایش آگاهی عمومی، و ارتقای سلامت جامعه، نقش مهمی در انتخاب‌های آگاهانه‌تر و سالم‌تر ایفا می‌کند.

## ۲-۱- اهداف پروژه

پروژه حاضر با هدف ارتقای سطح آگاهی و توانمندسازی مصرف‌کنندگان در انتخاب محصولات غذایی مناسب و سالم‌تر طراحی شده است. برای دستیابی به این هدف کلی، اهداف جزئی‌تری نیز تعریف شده‌اند که در ادامه به تفصیل به آنها پرداخته می‌شود:

### 1-2-1- افزایش آگاهی مصرف‌کنندگان از محتوای تغذیه‌ای محصولات غذایی

هدف اصلی این پروژه فراهم کردن دسترسی آسان و سریع مصرف‌کنندگان به اطلاعات تغذیه‌ای محصولات غذایی است. با استفاده از این برنامه، کاربران می‌توانند به‌راحتی میزان انرژی، قند، نمک، چربی‌ها، اسیدهای چرب ترانس و سایر مواد مغذی را مشاهده و تحلیل کنند. این اطلاعات به مصرف‌کنندگان کمک می‌کند تا با شناخت بهتری از محتوای محصولات، تصمیمات بهتری برای سلامتی خود بگیرند.

### 1-2-2- تسهیل فرآیند انتخاب محصولات غذایی سالم و مناسب

یکی دیگر از اهداف مهم این پروژه، ساده‌سازی فرآیند انتخاب محصولات غذایی سالم‌تر است. با استفاده از تکنولوژی پردازش تصویر و بینایی ماشین، برنامه قادر است اطلاعات پیچیده و گاه ناخوانای برچسب‌های مواد غذایی را به سرعت و به صورتی قابل فهم در اختیار کاربران قرار دهد. این امر به کاربران امکان می‌دهد تا در کوتاه‌ترین زمان ممکن، محصولاتی را انتخاب کنند که با نیازها و اهداف سلامتی آنها هماهنگ باشد.

### 1-2-3- ارائه اطلاعات فنی و بهداشتی محصولات

علاوه بر اطلاعات تغذیه‌ای، این پروژه به کاربران اجازه می‌دهد تا با بررسی شماره پروانه بهداشتی ساخت محصول، به اطلاعات فنی مهمی نظیر نام تجاری محصول، نام کارخانه تولیدکننده، تاریخ دریافت و انقضا پروانه بهداشتی، و آدرس کارخانه دسترسی پیدا کنند. این اطلاعات به کاربران کمک می‌کند تا از کیفیت و ایمنی محصولات مطمئن شوند و انتخاب‌های آگاهانه‌تری داشته باشند.

### 1-2-4- افزایش شفافیت در صنعت غذا

یکی دیگر از اهداف این پروژه، افزایش شفافیت در بازار محصولات غذایی است. با ارائه اطلاعات دقیق و معتبر درباره محتوا و مشخصات فنی محصولات، این پروژه به تقویت اعتماد مصرف‌کنندگان نسبت به اطلاعات ارائه‌شده توسط تولیدکنندگان کمک می‌کند و به ایجاد بازاری شفاف‌تر و قابل‌اعتمادتر منجر می‌شود.

### 1-2-5- حمایت از تولیدکنندگان متعهد به کیفیت

این پروژه می‌تواند به تولیدکنندگان محصولات غذایی که به کیفیت و شفافیت اطلاعات محصولات خود اهمیت می‌دهند، کمک کند تا در بازار رقابتی جایگاه بهتری پیدا کنند. با ارائه ابزارهایی که به مصرف‌کنندگان امکان دسترسی به اطلاعات دقیق و قابل‌اعتماد را می‌دهد، این پروژه به ترویج محصولات باکیفیت و افزایش اعتماد مشتریان به برندهای معتبر کمک می‌کند.

### 1-2-6- بهبود سلامت عمومی جامعه

هدف نهایی این پروژه ارتقای سلامت عمومی جامعه است. با افزایش دسترسی مصرف‌کنندگان به اطلاعات تغذیه‌ای و فنی دقیق، این پروژه به کاهش مصرف مواد مضر مانند قند، نمک و چربی‌های ناسالم کمک می‌کند. این امر می‌تواند به کاهش بروز بیماری‌های مرتبط با تغذیه نادرست، نظیر چاقی، دیابت، و بیماری‌های قلبی و عروقی منجر شود و در نهایت سطح کلی سلامت جامعه را بهبود بخشد.

این اهداف در مجموع به تحقق یک محیط غذایی سالم‌تر و شفاف‌تر کمک کرده و مصرف‌کنندگان را قادر می‌سازد تا با اطمینان و آگاهی بیشتری محصولات غذایی خود را انتخاب کنند.

## ۳-۱- دامنه و محدوده پروژه

### 1-3-1- جمع‌آوری و آماده‌سازی داده‌ها

* + - * **مجموعه داده**: برای آموزش و ارزیابی مدل‌های مورد استفاده، مجموعه‌ای از **1000 تصویر** از محصولات غذایی مختلف از چندین فروشگاه جمع‌آوری شده است. این تصاویر به منظور تشخیص برچسب محصول، تشخیص نوشته‌های داخل برچسب، و خواندن متن‌ها، برچسب‌گذاری[[3]](#footnote-3) و آماده‌سازی شده‌اند. این مجموعه داده برای آموزش سه مدل که در فصل‌های بعد به آنها اشاره خواهیم کرد، استفاده خواهدشد. در اینجا، استفاده این تصاویر و برپسب‌گذاری آنها به ترتیب آمده است.

1. **جمع آوری تصاویر از فروشگاه‌های مختلف**
2. **بازبینی تصاویر و تمیز کردن مجموعه داده**
3. **برچسب‌گذاری مجموعه داده برای مدل تشخیص‌دهنده[[4]](#footnote-4) برچسب و یافتن آن از داخل تصویر**
4. **برچسب‌گذاری مجموعه داده برای مدل تشخیص‌دهنده کلمات**
5. **برچسب‌ گذاری مجموعه داده برای مدل خواننده کلمات (OCR)**

برای این برچسب‌گذاری‌ها، از محیط‌های مختلفی استفاده شده که در فصل بعد آنها را معرفی می‌کنیم.

* + - * **برچسب‌گذاری و آماده‌سازی داده‌ها**: تصاویر به سه بخش اصلی لیبل‌گذاری شده‌اند:

1. **تشخیص برچسب محصول**: شناسایی موقعیت برچسب روی محصول.
2. **تشخیص نوشته‌های داخل برچسب**: شناسایی و تعیین مناطق متنی داخل برچسب.
3. **خواندن نوشته‌ها**: استخراج و تفسیر متن‌های موجود در برچسب.

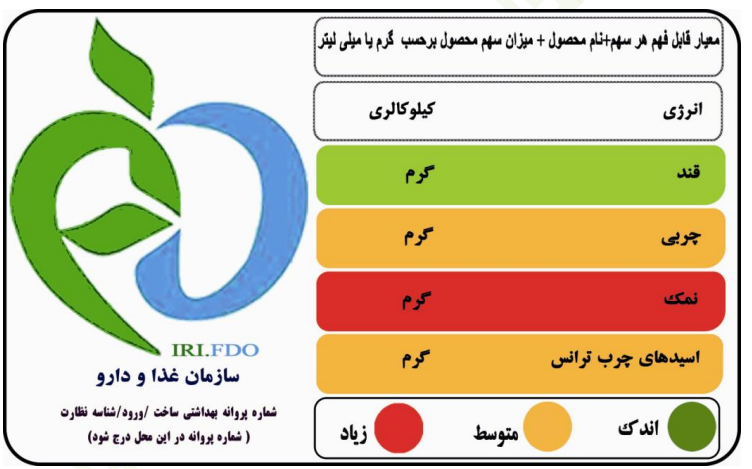
* **آموزش مدل‌ها[[5]](#footnote-5)**: مدل‌های یادگیری ماشین و پردازش تصویر با استفاده از این داده‌ها آموزش دیده‌اند تا به دقت بالا در شناسایی و استخراج اطلاعات برسند. نتیجه این آموزش‌ها و دقت مدل‌ها در فصل‌های بعد به نمایش گذاشته خواهد شد.

### 1-3-2- عملکرد برنامه در محیط واقعی

* **پویش و تحلیل برچسب‌ها**: برنامه طراحی شده است تا در محیط‌های فروشگاهی و با استفاده از دوربین تلفن همراه، برچسب‌های مختلف را پویش کرده و اطلاعات را به صورت دقیق استخراج کند.
* **کاربری و رابط کاربری**: رابط کاربری برنامه فعلا به گونه ای نیست که کاربران بتوانند برروی تلفن همراه خود و یا از طریق سایت به آن دسترسی داشته باشند.

### 1-3-3- دامنه پروژه

* **قالب استاندارد برچسب ارزش مواد غذایی**: این برنامه توانایی خواندن اکثر برچسب‌های موجود در فروشگاه‌ها را دارد اما قالب استانداردی که اداره کل نظارت و ارزیابی فراوده های خوراکی، آرایشی و بهداشتی سازمان غذاو دارو تایین کرده[1]، به عنوان قالب و چارچوب استاندارد این پروژه می‌باشد که به شرح زیر است.



1- برچسب استاندارد مد نظر در این پروژه

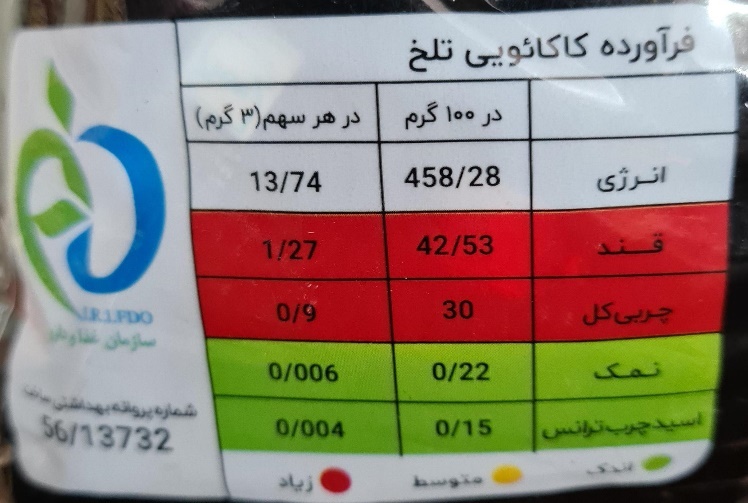
مواردی که در این برچسب به آن باید اشاره کرد به صورت زیر است.

* در بالاترین قسمت این برچسب، باید معیار قابل فهم هر سهم (برای مثال: یک قاشق غذاخوری از محصول یا هر 2 عدد از محصول)، نام محصول، و میزان سهم محصول برحسب گرم یا میلی‌لیتر قرار گیرد. در این قسمت، فقط بخش آخر که میزان سهم محصول است از برچسب خوانده می‌شود.
* در قسمت دوم برچسب، انرژی ای که در هر سهم از محصول موجود است و در قسمت اول ذکر شد، نوشته می‌شود. معیار اندازه گیری این قسمت، همیشه کیلوکالری (Kcal) است و اعداد، قبل از این معیار اندازه گیری قرار می‌گیرند.
* بعد از این دو قسمت، چهار قسمت رنگی وجود دارد که هرکدام یک ماده مغزی را نشان‌ می‌دهد. ترتیبی که این چهار ماده در برچسب دارند به ترتیب: قند، چربی، نمک و اسید‌های چرب ترانس است. معیار اندازه گیری این چهار ماده، گرم است و مانند انرژی، اعداد قبل از معیار قرار می‌گیرند.
* شماره پروانه بهداشتی پروانه ساخت نیز یکی دیگر از مواردیست که باید پویش شود و محل قرار گیری آن، پایین در سمت چپ است.

سایر مواردی که در برچسب هستند، در این پروژه کاربردی ندارند و پویش نخواهند شد.

### 1-3-4- خارج از دامنه پروژه

* **محصولات غیر بسته‌بندی**: پروژه محدود به محصولات غذایی بسته‌بندی‌شده با برچسب‌های واضح و قابل خواندن است و شامل مواد غذایی فاقد برچسب، برچسب غیر استاندارد سازمان غذا و دارو، برچسب‌هایی که به طور کامل و صحیح عکس‌برداری نشده و یا برچسب‌های بسیار آسیب‌دیده نمی‌شود. چند نمونه از این برچسب‌ها را می‌توان در زیر مشاهده کرد.



2- نمونه ای از برچسب غیر استاندارد



3- نمونه ای از برچسب با عکس برداری نامناسب



4- نمونه ای از برچسب با اعداد ناخوانا



5- نمونه ای از برچسب مخدوش

* **پشتیبانی از ویژگی‌های خاص**: برنامه به طور خاص برای استخراج و نمایش اطلاعات تغذیه‌ای و فنی طراحی شده است و ویژگی‌های خاصی مانند تجزیه و تحلیل تغذیه‌ای پیشرفته یا مشاوره‌های بهداشتی را شامل نمی‌شود. همچنین رنگ‌ها در برچسب ارزش مواد غذایی، ارزیابی نمی‌شوند.

### 1-4- چالش ها

* سیستم‌های OCR معمولاً بر روی متن‌های معمولی آموزش داده می‌شوند که توسط جملات و پاراگراف‌ها تشکیل می‌شوند، اما برچسب‌های تغذیه‌ای جدولی هستند و تاکید زیادی بر اعداد دارند. ثانیاً، این نمودارها از نظر رنگ‌ها، روشنایی، فونت و اندازه متن، الگوها، بافت‌ها و غیره بسیار متغیر هستند. که توسعه یک سیستم عمومی که قادر به تعمیم همه این شرایط باشد، بسیار چالش برانگیز است.
* تصاویر گرفته‌شده از برچسب‌ها ممکن است شامل نویزهای پس‌زمینه‌ای مانند بازتاب نور، سایه، یا حتی سایر محصولات مجاور باشند که می‌توانند باعث کاهش دقت در تشخیص متن شوند. فیلتر کردن این نویزها بدون از دست دادن اطلاعات مهم از جمله چالش‌های اساسی است.
* برخی برچسب‌ها دارای اطلاعات بسیار فشرده هستند که در یک فضای کوچک قرار داده شده‌اند. تشخیص دقیق این متون فشرده که ممکن است از فونت‌های بسیار کوچک استفاده کنند، چالشی برای OCR به حساب می‌آید.
* در صورتی که برنامه باید در زمان واقعی عمل کند، یعنی کاربر بتواند به محض گرفتن عکس از برچسب، اطلاعات را مشاهده کند، چالش‌هایی مانند بهینه‌سازی سرعت پردازش و کاهش تاخیر زمانی مطرح می‌شوند. این موضوع به خصوص در دستگاه‌های موبایل با منابع محدود مانند حافظه و پردازنده، چالش‌برانگیز است.

این چالش‌ها نه تنها اهمیت پیچیدگی پروژه را نشان می‌دهند، بلکه نیاز به استفاده از رویکردهای نوآورانه و بهینه‌سازی‌های مداوم را برای دستیابی به نتایج مطلوب تأکید می‌کنند.

# فصل دوم

# ادبیات پژوهش

## ۱-۲- مقدمه

در این فصل، به توضیح مفاهیم کلیدی مطرح‌شده در این پایان‌نامه پرداخته و به اختصار هر یک را بررسی می‌کنیم. در انجام این پژوهش، از ابزارها، روش‌ها و الگوریتم‌های مختلفی در حوزه های مختلف استفاده شده است که در این فصل به کاربرد هر یک از آن‌ها در این پروژه اشاره خواهیم کرد.

مطالب این فصل به سه بخش اصلی تقسیم شده است: ادبیات داده، ادبیات مدل، و ادبیات ارزیابی. در هر بخش، مباحث مرتبط با آن حوزه به طور جداگانه و در ارتباط با پروژه حاضر مورد بررسی قرار می‌گیرند.

**2-۲- ادبیات داده**

بخش مهمی از این پایان‌نامه به ایجاد و توسعه مجموعه‌داده‌ای برای برچسب‌های محصولات داخلی اختصاص یافته است. در این مسیر، از ابزارها و کتابخانه‌های متعددی استفاده شده که هر یک نقش کلیدی در فرآیند برچسب زدن، استخراج و پیش‌پردازش داده‌ها ایفا کرده‌اند. این ابزارها و کتابخانه‌ها نه تنها برای جمع‌آوری داده‌ها، بلکه برای پاک‌سازی، اصلاح، و آماده‌سازی آنها جهت استفاده در مدل‌های مختلف به کار گرفته شده‌اند.

در ادامه، برخی از این ابزارها و کتابخانه‌ها به تفصیل معرفی خواهند شد و نقش و کاربرد دقیق هر یک از آن‌ها در این پایان‌نامه مورد بررسی قرار می‌گیرد. همچنین، به نحوه استفاده از این ابزارها در مراحل مختلف پروژه، از جمله استخراج داده‌های خام، پردازش اولیه، و آماده‌سازی نهایی برای تجزیه و تحلیل و مدل‌سازی، پرداخته می‌شود. هدف این بخش ارائه یک دیدگاه جامع از ابزارها و تکنیک‌هایی است که به پیشبرد تحقیق کمک کرده‌اند و نقش آنها در ایجاد یک مجموعه‌داده معتبر و قابل‌اعتماد برای زبان فارسی را تبیین می‌کند.

**1-2-1- فرایند برچسب گذاری تصویر[[6]](#footnote-6)**

فرایند برچسب‌گذاری تصویر یک مرحله مهم در آماده‌سازی داده‌ها برای پروژه‌های یادگیری ماشین، به‌ویژه در حوزه بینایی ماشین و تشخیص اشیاء، است. این فرایند شامل شناسایی و برچسب‌گذاری نواحی خاص در تصاویر است که مدل‌های یادگیری ماشین بر اساس آن آموزش می‌بینند. در ادامه، مراحل اصلی این فرایند را توضیح می‌دهم:

1. **جمع‌آوری تصاویر**

اولین مرحله شامل جمع‌آوری مجموعه‌ای از تصاویر مرتبط با هدف پروژه است. این تصاویر می‌توانند از منابع مختلفی مانند دوربین‌ها، پایگاه‌های داده عمومی، اینترنت، و غیره جمع‌آوری شوند.

1. **انتخاب ابزار برچسب‌گذاری**

ابزارهای فراوانی برای برچسب‌گذاری وجود دارد که چند نمونه در قسمت بعدی توضیح داده شده است.

1. **تعریف برچسب‌ها**

قبل از شروع برچسب‌گذاری، باید دسته‌بندی‌ها یا کلاس‌هایی که قرار است برچسب‌گذاری شوند، تعریف شود. برای مثال، در این پروژه، برچسب‌ها شامل برچسب ارزش غذایی، اعداد داخل برچسب، مواد مغذی، واحدهای اندازه گیری و پروانه بهداشتی ساخت می‌باشند.

1. **برچسب‌گذاری نواحی**

در این مرحله، کاربر با استفاده از ابزار برچسب‌گذاری، نواحی مربوط به هر دسته‌بندی را در تصاویر انتخاب می‌کند. این کار می‌تواند شامل ترسیم کادرهای محدودکننده[[7]](#footnote-7)، ماسک‌های تقسیم‌بندی[[8]](#footnote-8) یا نقاط کلیدی [[9]](#footnote-9)باشد که به شناسایی دقیق نواحی مورد نظر کمک می‌کند.

1. **بازبینی و تأیید**

پس از برچسب‌گذاری اولیه، نواحی برچسب‌گذاری‌شده بازبینی می‌شوند تا اطمینان حاصل شود که برچسب‌ها به درستی و با دقت اعمال شده‌اند. در پروژه‌های بزرگتر، این مرحله ممکن است توسط افراد مختلف یا با استفاده از الگوریتم‌های خودکار انجام شود.

1. **ذخیره‌سازی داده‌های برچسب‌گذاری‌شده**

پس از تأیید، برچسب‌ها به همراه تصویر در قالب‌های استاندارد مانند ایکس‌ام‌ال[[10]](#footnote-10)، جیسان[[11]](#footnote-11)، یا فایل‌های متنی ذخیره می‌شوند. این داده‌ها سپس برای آموزش مدل‌های یادگیری ماشین استفاده می‌شوند.

1. **آماده‌سازی برای آموزش مدل**

در نهایت، داده‌های برچسب‌گذاری‌شده برای آموزش مدل آماده می‌شوند. این مرحله شامل تقسیم‌بندی داده‌ها به مجموعه‌های آموزشی، اعتبارسنجی و آزمایش است که طبق ساختاری که برای آموزش هر مدل مشخص شده است، مرتب می‌شوند.

فرایند برچسب‌گذاری تصویر یک گام اساسی برای اطمینان از دقت و عملکرد بالای مدل‌های یادگیری ماشین در پروژه‌های بینایی ماشین است. کیفیت داده‌های برچسب‌گذاری‌شده به طور مستقیم بر عملکرد نهایی مدل تأثیر می‌گذارد.

**1-2-2- پلتفرم[[12]](#footnote-12) روبوفلو[[13]](#footnote-13)**

روبوفلو یک پلتفرم جامع برای توسعه پروژه‌های بینایی ماشین است که به ویژه در بخش برچسب‌گذاری تصاویر بسیار قدرتمند است. این پلتفرم ابزارهای متنوعی برای برچسب‌گذاری دستی، نیمه‌خودکار، و خودکار ارائه می‌دهد که کاربران می‌توانند با آن‌ها به سادگی نواحی موردنظر در تصاویر را مشخص کنند. RoboFlow از انواع مدل‌های بینایی ماشین مانند تشخیص اشیاء، تقسیم‌بندی، و طبقه‌بندی پشتیبانی می‌کند.

همچنین، این پلتفرم امکاناتی برای مدیریت نسخه‌ها، همکاری تیمی، و بهینه‌سازی داده‌ها فراهم می‌کند. به لطف یکپارچگی با فریم‌ورک‌های یادگیری ماشین محبوب، داده‌های برچسب‌گذاری‌شده می‌توانند مستقیماً برای آموزش مدل‌ها استفاده شوند، که باعث افزایش دقت و سرعت فرآیند توسعه مدل‌های بینایی ماشین می‌شود.[2].

**1-2-3- پلتفرم لیبل استودیو[[14]](#footnote-14)**

لیبل استودیو یک پلتفرم متن‌باز**[[15]](#footnote-15)**  برای برچسب‌گذاری داده‌ها است که از انواع مختلف داده‌ها از جمله تصاویر، ویدئوها، متن‌ها و فایل‌های صوتی پشتیبانی می‌کند. این ابزار به کاربران امکان می‌دهد تا به راحتی پروژه‌های برچسب‌گذاری خود را مدیریت کرده و داده‌هایشان را برای مدل‌های یادگیری ماشین آماده کنند. با قابلیت‌های سفارشی‌سازی بالا، کاربران می‌توانند رابط کاربری و ابزارهای برچسب‌گذاری را متناسب با نیازهای خاص پروژه‌های خود تنظیم کنند.

یکی از ویژگی‌های برجسته این پلتفرم، امکان همکاری همزمان چندین کاربر در برچسب‌گذاری داده‌ها است که باعث افزایش سرعت و دقت کار می‌شود. مدیریت پروژه‌های برچسب‌گذاری نیز با این پلتفرم بسیار ساده‌تر می‌شود، به طوری که کاربران می‌توانند فرآیندهای کاری، تخصیص وظایف و مراحل بازبینی را به طور دقیق کنترل کنند. این ابزار همچنین با فریم‌ورک‌های یادگیری ماشین یکپارچه می‌شود، که داده‌های برچسب‌گذاری‌شده را مستقیماً برای آموزش مدل‌ها قابل استفاده می‌کند.

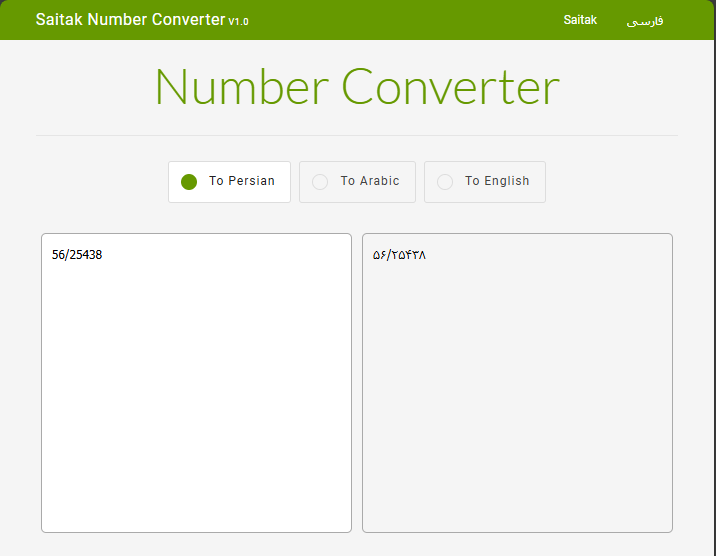
علاوه بر این، از افزونه‌ها و ای‌پی‌آی‌[[16]](#footnote-16)های متنوعی پشتیبانی می‌کند که امکان یکپارچگی با سایر سیستم‌ها و گسترش قابلیت‌های پلتفرم را فراهم می‌سازد. این ویژگی‌ها به کاربران اجازه می‌دهد تا این پلتفرم را با ابزارهای موجود خود ترکیب کرده و از آن در محیط‌های مختلف استفاده کنند[3].



6- محیط لیبل استودیو و فرایند برچسب‌کذاری

**1-2-4- وب‌سایت سایتاک[[17]](#footnote-17)**

این وبسایت، بستری برای تبدیل کاراکتر اعداد انگلیسی به کاراکتر اعداد فارسی فراهم‌ می‌کند. این تبدیل از این جهت مفید است که در پلتفرم‌های برچسب گذاری، اعداد فارسی با استفاده از صفحه‌کلید، در دسترس نیستند و برای آموزش مدل‌ها، باید اعداد فارسی بر روی برچسب را، با استفاده از کاراکتر‌های فارسی نوشت[4].



7- محیط وب‌سایت سایتاک

**1-3-1- پلتفرم اوپن‌سی‌وی[[18]](#footnote-18)**

این کتابخانه یک کتابخانه متن‌باز است که به طور گسترده برای برنامه‌های بینایی ماشین و پردازش تصویر استفاده می‌شود. این کتابخانه ابتدا توسط اینتل[[19]](#footnote-19) توسعه یافت و اکنون توسط یک جامعه کاربری بزرگ نگهداری می‌شود. اوپن‌سی‌وی شامل بیش از 2500 الگوریتم بهینه‌سازی شده است که برای تشخیص و ردیابی اشیاء، شناسایی چهره، تحلیل حرکات، و پردازش تصاویر مورد استفاده قرار می‌گیرند. یکی از ویژگی‌های برجسته این کتابخانه این است که از چندین زبان برنامه‌نویسی پشتیبانی می‌کند و همچنین با سیستم‌عامل‌های مختلفی مانند ویندوز، لینوکس[[20]](#footnote-20)، و مک‌او‌اس[[21]](#footnote-21) سازگار است.

اوپن‌سی‌وی در طیف وسیعی از کاربردها مانند بینایی ماشین، رباتیک، تحلیل ویدئو، و واقعیت افزوده استفاده می‌شود. از جمله قابلیت‌های این کتابخانه می‌توان به تشخیص لبه‌ها، فیلترگذاری تصاویر، شناسایی ویژگی‌ها، و پردازش ویدئو اشاره کرد. اوپن‌سی‌وی به دلیل سرعت و کارایی بالا در پردازش تصویر، در پروژه‌های تحقیقاتی و تجاری به طور گسترده‌ای مورد استفاده قرار می‌گیرد. این کتابخانه به توسعه‌دهندگان امکان می‌دهد تا با استفاده از ابزارهای قدرتمند و الگوریتم‌های از پیش ساخته شده، برنامه‌های پیچیده بینایی ماشین را به سرعت پیاده‌سازی کنند.

در این پروژه از اوپن‌سی‌وی عموماً برای خواندن عکس‌ها، نمایش اطلاعات خوانده شده بر روی هر عکس، بازبینی دقت و عملکرد مدل‌ها و ذخیره عکس‌ها استفاده شده‌است.

**1-3-1- کتابخانه پیل[[22]](#footnote-22)**

پیل یک کتابخانه قدرتمند و کاربردی برای پردازش تصاویر در پایتون است. این کتابخانه امکان باز کردن، تغییر، و ذخیره انواع مختلف تصاویر را فراهم می‌کند و از فرمت‌های تصویری متعددی مانند جی‌پی‌جی[[23]](#footnote-23)، پی‌ان‌جی[[24]](#footnote-24)، گیف[[25]](#footnote-25)، و بی‌ام‌پی[[26]](#footnote-26) پشتیبانی می‌کند. با استفاده از پیل، توسعه‌دهندگان می‌توانند عملیات مختلفی مانند تغییر اندازه، برش، فیلترگذاری، و تبدیل فرمت تصاویر را به سادگی انجام دهند. اگرچه توسعه رسمی پیل متوقف شده است، اما نسخه‌ای به‌روزتر و توسعه‌یافته به نام پیلو[[27]](#footnote-27) به عنوان جایگزین آن مورد استفاده قرار می‌گیرد و به طور گسترده‌ای در پروژه‌های پردازش تصویر استفاده می‌شود.

مشکل اصلی اوپین‌سی‌وی، نمایش کاراکترهای فارسی بر روی عکس است که باعث به هم ریختگی در حرف و ارقام می‌شود. این مشکل با استفاده از کتابخانه پیل برطرف شده و حروف و ارقام به ترتیب درست، بر روی عکس درج می‌شوند.

**1-3-1- کتابخانه نامپای[[28]](#footnote-28)**

نامپای، یک کتابخانه اساسی و محبوب در زبان برنامه‌نویسی پایتون است که برای محاسبات علمی و عددی استفاده می‌شود. نامپای به‌ویژه به دلیل ارائه آرایه‌های چندبعدی[[29]](#footnote-29) با عملکرد بالا شناخته شده است. این آرایه‌ها به توسعه‌دهندگان اجازه می‌دهند تا عملیات‌های پیچیده ریاضی و محاسباتی را به صورت کارآمد و سریع انجام دهند. برخلاف لیست‌های پایتون، آرایه‌های نامپای دارای ابعاد ثابت و نوع داده یکنواخت هستند که این ویژگی‌ها باعث بهبود سرعت و کاهش مصرف حافظه می‌شود. علاوه بر این، نامپیا شامل تعداد زیادی از توابع ریاضی برای انجام محاسبات مختلف مانند جبر خطی، تبدیل فوریه، و تولید اعداد تصادفی است.

نامپای به عنوان یکی از کتابخانه‌های پایه‌ای در عملیات داده‌کاوی[[30]](#footnote-30) و یادگیری ماشین پایتون عمل می‌کند و بسیاری از کتابخانه‌های دیگر به طور مستقیم بر پایه نامپای ساخته شده‌اند. این کتابخانه به دلیل قابلیت‌های گسترده و عملکرد بالا، به‌ویژه در پردازش داده‌های بزرگ و انجام تحلیل‌های عددی در پروژه‌های علمی و مهندسی مورد استفاده قرار می‌گیرد. به عنوان مثال، توسعه‌دهندگان می‌توانند با استفاده از نامپای ماتریس‌ها و بردارها را به سادگی مدیریت کنند و محاسبات پیچیده‌ای مانند ضرب ماتریسی، برون‌یابی، و انتگرال‌گیری عددی را انجام دهند.

**1-3-1- کتابخانه ره[[31]](#footnote-31)**

ره کتابخانه‌ای در پایتون است که برای کار با عبارات منظم[[32]](#footnote-32) طراحی شده است. این کتابخانه امکاناتی را برای جستجو، جایگزینی، و تطبیق الگوهای متنی فراهم می‌کند و به توسعه‌دهندگان اجازه می‌دهد تا با استفاده از الگوهای پیشرفته، به سرعت بخش‌های خاصی از متن‌ها را شناسایی و پردازش کنند. با این کتابخانه، می‌توان عملیات‌های مختلفی مانند بررسی وجود الگو، تقسیم متن بر اساس الگو، و جایگزینی متن‌های مطابق با الگو را به سادگی انجام داد. این ابزار به ویژه در پردازش داده‌ها، تحلیل متن، و اعتبارسنجی ورودی‌های کاربر کاربرد زیادی دارد.

**1-3-1- کتابخانه‌های اصلاح قالب[[33]](#footnote-33)**

عربیک‌ریشیپر[[34]](#footnote-34) و بیدی[[35]](#footnote-35) دو کتابخانه پایتون هستند که برای پردازش و نمایش صحیح متن‌های عربی و دیگر زبان‌های راست به چپ[[36]](#footnote-36) مانند فارسی استفاده می‌شوند. عربیک‌ریشیپر وظیفه تغییر شکل حروف عربی را به گونه‌ای دارد که آن‌ها به درستی به هم متصل شوند، زیرا حروف در این زبان‌ها بسته به موقعیتشان در کلمه، اشکال مختلفی دارند. از سوی دیگر، بیدی برای ترتیب‌گذاری صحیح متن‌های دوزبانه (شامل حروف از راست به چپ و چپ به راست) به کار می‌رود، به‌ویژه وقتی که اعداد یا کلمات لاتین در متن عربی یا فارسی قرار می‌گیرند. این دو کتابخانه معمولاً در کنار هم استفاده می‌شوند تا متن‌های عربی و فارسی به درستی در رابط‌های گرافیکی یا تصاویر نمایش داده شوند. این دو کتابخانه همراه با کتابخانه پیل، برای قرار دادن ارقام و حروف فارسی بر روی عکس، استفاده شده‌اند.

**1-3-1- کتابخانه‌‌ی جیسان**

جیسان کتابخانه‌ای در پایتون است که برای کار با داده‌های فرمت جیسان طراحی شده است. این کتابخانه امکاناتی را برای تبدیل داده‌های پایتون به فرمت جیسان و بالعکس فراهم می‌کند، به طوری که می‌توان به راحتی داده‌ها را بین پایتون و دیگر زبان‌های برنامه‌نویسی یا سیستم‌های مبتنی بر جیسان مبادله کرد. در این پروژه از جیسان برای خواندن داده‌های برچسب‌گذاری شده که در پلتفرم‌های مختلف آماده شده، استفاده می‌شود.

**3-۲- ادبیات مدل**

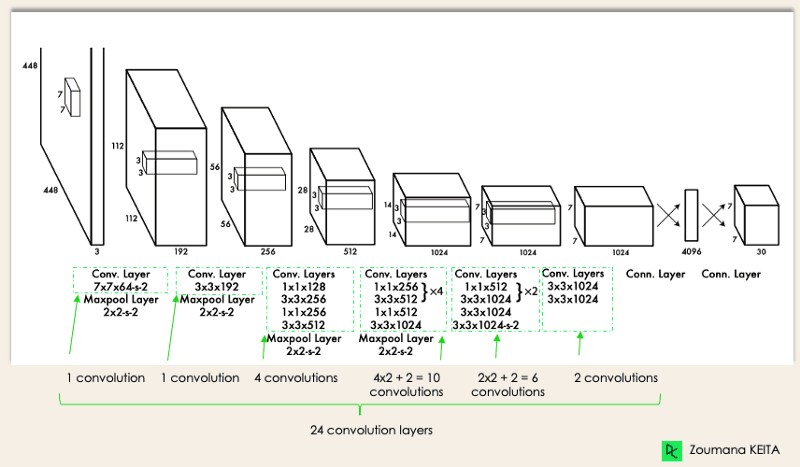
در این پایان‌نامه، برای اینکه بتوانیم حروف، کلمات و اعداد نوشته شده بر روی برچسب را بخوانیم، نیاز بود که مدل‌هایی را با استفاده از آنچه که گفته شد، آموزش دهیم و از آن استفاده کنیم. برای آموزش این مدل‌ها از ابزارها و کتابخانه‌هایی استفاده شد که در این بخش به معرفی آنها می‌پردازیم. همچنین هر آن چیزی که به سرعت بخشیدن و بهبود این فرایند کمک کرد را نیز معرفی خواهیم کرد.

**1-3-1- پلتفرم آلترالیتیکس[[37]](#footnote-37)**

آلترالیتیکسیک کتابخانه پایتون است که به ویژه برای کار با مدل‌های یادگیری عمیق در حوزه بینایی ماشین توسعه یافته است. این کتابخانه به خاطر پیاده‌سازی ساده و کاربردی از مدل‌های پیشرفته‌ی بینایی ماشین مانند یولو[[38]](#footnote-38) مشهور است. این کتابخانه، به توسعه‌دهندگان این امکان را می‌دهد که مدل‌های تشخیص اشیاء و دیگر وظایف بینایی ماشین را با استفاده از پیکربندی‌های ساده و استاندارد به سرعت آموزش دهند و اجرا کنند.

**1-3-1- الگوریتم یولو**

یولو یک الگوریتم پیشرفته برای تشخیص اشیاء در تصاویر و ویدئوها است که به دلیل سرعت و دقت بالای خود شناخته شده است. برخلاف روش‌های سنتی که تصویر را به بخش‌های کوچک‌تر تقسیم کرده و هر بخش را به صورت جداگانه تحلیل می‌کنند، یولو کل تصویر را به یکباره پردازش کرده و به شناسایی و مکان‌یابی اشیاء در آن می‌پردازد. این الگوریتم با استفاده از شبکه‌های عصبی کانولوشنی به سرعت اشیاء مختلف را در تصویر شناسایی کرده و جعبه‌های محدود کننده و برچسب‌های مربوط به هر شیء را پیش‌بینی می‌کند. یولو به دلیل عملکرد بالا و توانایی پردازش در زمان واقعی، در کاربردهایی مانند نظارت ویدئویی، رباتیک، و تحلیل تصاویر پزشکی بسیار مورد استفاده قرار می‌گیرد. ساختار کلی این مدل، در شکل زیر به صورت واضح قابل مشاهده است.



8- ساختار کلی مدل یولو

در این پروژه از یولو ورژن 8 برای تشخیص برچسب ارزش مواد غذایی، استفاده شده است. این نسخه، به عنوان نسخه جدید و بهبود یافته الگوریتم یولو، به دلیل مزایای متعددی که ارائه می‌دهد، انتخاب مناسبی برای پروژه‌های بینایی ماشین است. این نسخه با بهینه‌سازی‌های الگوریتمی و معماری پیشرفته‌تر، دقت بالاتری در تشخیص اشیاء و سرعت پردازش سریع‌تری نسبت به نسخه‌های قبلی ارائه می‌دهد. به خاطر توانایی‌های بهبود یافته در شناسایی و مکان‌یابی دقیق اشیاء، قابلیت کار در زمان واقعی، و مقیاس‌پذیری برای پردازش داده‌های بزرگ و متنوع، به ویژه در پروژه‌هایی که نیاز به پردازش سریع و دقیق تصاویر دارند، بسیار مناسب است.

**1-3-1- پلتفرم پایتورچ[[39]](#footnote-39)**

پایتورچ یک کتابخانه متن‌باز و محبوب برای یادگیری ماشین و یادگیری عمیق است که توسط فیس‌بوک[[40]](#footnote-40) توسعه یافته است. این کتابخانه به دلیل سهولت استفاده و انعطاف‌پذیری بالا، به ویژه در میان پژوهشگران و توسعه‌دهندگان یادگیری عمیق محبوبیت زیادی دارد. پایتورچ از تنسور[[41]](#footnote-41) برای انجام محاسبات عددی استفاده می‌کند و با گراف‌[[42]](#footnote-42) محاسباتی پویا، امکان طراحی و آزمایش مدل‌های پیچیده را فراهم می‌سازد. این ویژگی‌ها به توسعه‌دهندگان کمک می‌کند تا به سرعت مدل‌های خود را توسعه داده و آزمایش کنند.

پایتورچ به دلیل سازگاری با ابزارها و کتابخانه‌های دیگر مانند نامپای[[43]](#footnote-43) و سایکیت‌لرن[[44]](#footnote-44) و همچنین پشتیبانی گسترده از سوی جامعه کاربری، یک انتخاب محبوب برای پروژه‌های یادگیری عمیق در حوزه‌هایی مانند پردازش تصویر و پردازش زبان طبیعی است. این کتابخانه به کاربران در سطوح مختلف اجازه می‌دهد تا به راحتی مدل‌های یادگیری ماشین را پیاده‌سازی و آموزش دهند، و در نتیجه در پروژه‌های تحقیقاتی و تجاری بسیار مورد استفاده قرار می‌گیرد.

در اين پايان‌نامه نيز براي پياده‌سازي مدل‌ها و استفاده از آن‌ها، از اين کتابخانه استفاده‌ شده است و مدل‌هایی که در این پروژه نام برده خواهند شد، در این فرمت ذخیره شده اند.

**1-3-1- پلتفرم محاسبات موازی کودا[[45]](#footnote-45)**

کودا یک پلتفرم محاسباتی موازی و رابط برنامه‌نویسی کاربردی است که توسط انویدیا[[46]](#footnote-46) توسعه یافته است. این فناوری به توسعه‌دهندگان اجازه می‌دهد تا از قدرت پردازشی بالای پردازنده‌های گرافیکی[[47]](#footnote-47) برای انجام محاسبات موازی و تسریع پردازش‌ها استفاده کنند. با بهره‌گیری از هزاران هسته پردازشی موجود در کارت گرافیک، کودا می‌تواند زمان محاسبات پیچیده را به طور قابل توجهی کاهش دهد، و این امر به ویژه در زمینه‌هایی مانند یادگیری ماشین و شبیه‌سازی‌های علمی کاربرد دارد.

کودا با زبان‌های برنامه‌نویسی مانند سی[[48]](#footnote-48)، سی‌پلاس‌پلاس[[49]](#footnote-49) و پایتون یکپارچه شده است و بسیاری از کتابخانه‌های یادگیری عمیق مانند تنسورفلو[[50]](#footnote-50) و پایتورچ از آن برای تسریع آموزش مدل‌ها استفاده می‌کنند. این تکنولوژی به توسعه‌دهندگان امکان می‌دهد تا از کارت گرافیک برای انجام محاسبات پیچیده و زمان‌بر به صورت کارآمدتر استفاده کنند، که در نهایت به بهبود عملکرد در کاربردهای مختلفی مانند تحلیل داده‌های بزرگ، رندرینگ[[51]](#footnote-51) گرافیکی و یادگیری عمیق منجر می‌شود.

**1-3-1- پلتفرم کگل[[52]](#footnote-52)**

کگل یک پلتفرم جامع برای تحلیل داده‌ها و مسابقات یادگیری ماشین است که به محققان، دانشمندان داده و توسعه‌دهندگان این امکان را می‌دهد که به صورت آنلاین به مجموعه‌های داده، ابزارهای تحلیلی، و جوامع فعال دسترسی پیدا کنند. کاربران کگل می‌توانند در مسابقات داده‌ای شرکت کنند، مدل‌های یادگیری ماشین را پیاده‌سازی کرده و با دیگران به اشتراک بگذارند. این پلتفرم به ویژه برای آموزش و آزمون مدل‌های مختلف مفید است و از طریق ارائه داده‌های معتبر و ابزارهای قدرتمند، به پیشبرد تحقیق و توسعه در زمینه‌های مختلف کمک می‌کند.

علاوه بر مسابقات، کگل یک محیط تعاملی برای جستجو و تحلیل داده‌ها ارائه می‌دهد که شامل نوت‌بوک[[53]](#footnote-53)‌ها و یک پایگاه داده بزرگ از مجموعه‌های داده عمومی است. کاربران می‌توانند با استفاده از زبان‌های برنامه‌نویسی مختلف مانند پایتون[[54]](#footnote-54) و آر[[55]](#footnote-55) به تحلیل داده‌ها بپردازند و کد‌ها و مدل‌های خود را بر روی سخت‌افزار‌‌های مختلف از جمله کارت گرافیک‌ها و پردازنده‌ها اجرا کنند. همچنین می‌توانند نتایج خود را با جامعه کگل به اشتراک بگذارند. این ویژگی‌ها به محققان و توسعه‌دهندگان این امکان را می‌دهد که به سرعت به داده‌های واقعی دسترسی پیدا کنند، مدل‌های خود را ارزیابی کنند و از بازخورد جامعه برای بهبود عملکرد مدل‌های خود بهره‌برداری کنند[5].

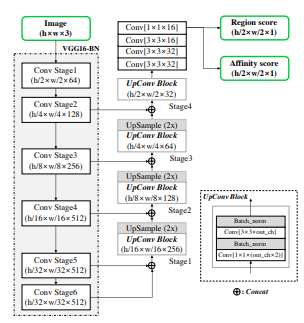
**1-3-2- پلتفرم ایزی‌او‌سی‌آر[[56]](#footnote-56)**

ایزی‌او‌سی‌آر یک کتابخانه منبع باز برای تشخیص و شناسایی متن از تصاویر است که با هدف ارائه یک راه‌حل ساده و سریع برای پردازش متن تصویری طراحی شده است. این کتابخانه به ویژه برای پروژه‌های بینایی ماشین که نیاز به استخراج متن از تصاویر دارند، مناسب است. ایزی‌او‌سی‌آر از مدل‌های پیش‌آموزش‌دیده استفاده می‌کند و به کاربران این امکان را می‌دهد که بدون نیاز به آموزش مدل‌های جدید، به راحتی متن‌ها را از تصاویر مختلف استخراج کنند. با پشتیبانی از زبان‌های مختلف و رابط کاربری ساده، این ابزار به توسعه‌دهندگان و محققان کمک می‌کند تا به سرعت و با دقت بالا متن‌های تصویری را شناسایی کنند.

این پلتفرم شامل دو مدل اصلی است. مدل تشخیص[[57]](#footnote-57) و مدل شناسایی[[58]](#footnote-58). مدل تشخیص مسئول شناسایی نواحی متنی در تصویر است و به تعیین مکان‌های دقیق متن‌ها کمک می‌کند. این مدل قادر است نواحی مختلف متن را به درستی تشخیص دهد، حتی در شرایط پیچیده مانند تغییرات در اندازه، فونت و زاویه متن. مدل شناسایی به دنبال آن، وظیفه شناسایی و استخراج محتوای متنی از نواحی مشخص‌شده را بر عهده دارد. این مدل متن‌های شناسایی شده را به صورت خوانا و قابل استفاده استخراج می‌کند، و به این ترتیب، مراحل پردازش تصویر و استخراج متن به طور کامل و دقیق انجام می‌شود[6]. در ادامه به معرفی دو مدلی که در این پلتفرم استفاده شده می‌پردازیم[6].

1. **کرفت[[59]](#footnote-59)**

مدل تشخیص متن در ایزی‌او‌سی‌آر کرفت است. کرفتیک مدل پیشرفته برای شناسایی مناطق متنی در تصاویر است که به ویژه برای تشخیص متن در تصاویری با پس‌زمینه پیچیده و نویزهای زیاد طراحی شده است. این مدل با استفاده از تکنیک‌های یادگیری عمیق، نواحی متنی را با دقت بالا شناسایی می‌کند و قادر است متن‌های کوچک و نازک را که توسط مدل‌های دیگر ممکن است نادیده گرفته شوند، شناسایی کند[7].



9- ساختار کرفت

1. **شناسایی عمیق متن[[60]](#footnote-60)**

یک مجموعه از مدل‌ها و ابزارها برای شناسایی و خواندن متن از تصاویر است که به منظور مقایسه عملکرد الگوریتم‌های مختلف شناسایی متن طراحی شده است. این ریپازیتوری شامل کدهایی برای آموزش و ارزیابی مدل‌های مختلف شناسایی متن و همچنین ابزارهایی برای مقایسه عملکرد آنها است[8].

## 4-3- ادبیات ارزیابی:

# فصل سوم

# پیشینه پژوهش

## ۱-3- مقدمه:

این فصل به بررسی تحقیقات و پروژه‌های پیشین در زمینه پردازش تصویرو تشخیص نوری حروف[[61]](#footnote-61) می‌پردازد. در این زمینه، تحقیقات زیادی صورت نگرفته است، از دلایل آن می‌توان به تفاوت بین چاپ برچسب ارزش مواد غذایی بر روی محصولات مختلف و استفاده از رنگ‌ها با کد‌های رنگی مختلف نام برد. چند نمونه تحقیق در این زمینه بر روی برچسب محصولات آمریکایی[[62]](#footnote-62) انجام شده و نمونه این پویشگر بر روی محصولات ایرانی، وجود ندارد.

این ادبیات، چارچوبی را فراهم می‌کند تا نشان دهد چگونه پروژه فعلی می‌تواند به پیشرفت دانش موجود در این زمینه و گسترش آن به زبان فارسی کمک کند و راهکارهایی نوآورانه برای حل مشکلات و بهبود تجربه کاربری ارائه دهد.

## ۲-3- تاریخچه و توسعه OCR:

تشخیص نوری حروف یکی از تکنولوژی‌های کلیدی در پردازش تصویر است که برای استخراج متون از تصاویر استفاده می‌شود. تکنولوژی OCR از دهه 1950 توسعه یافته و به مرور زمان با بهبودهای مداوم در دقت و کارایی روبرو شده است. از جمله پیشرفت‌های اخیر می‌توان به استفاده از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی مصنوعی اشاره کرد که دقت تشخیص حروف را به طور قابل توجهی افزایش داده‌اند. به عنوان مثال، الگوریتم‌هایی نظیر Tesseract که به صورت متن باز توسعه یافته‌اند، قادر به تشخیص متون در چندین زبان و با دقت بالایی هستند.

مدل‌های یادگیری عمیق، مانند شبکه‌های عصبی کانولوشنی[[63]](#footnote-63)و شبکه‌های بازگشتی (RNN)[[64]](#footnote-64)، به دلیل دقت بالا در تشخیص و طبقه‌بندی تصاویر، به طور گسترده‌ای در OCR مورد استفاده قرار گرفته‌اند. مدل‌های مدرنی نظیر MaskOCR و TransOCR از جدیدترین الگوریتم‌ها در این حوزه هستند که دقت‌های بسیار بالایی در تشخیص متن از تصاویر دارند. این مدل‌ها با استفاده از مجموعه داده‌های بزرگ و متنوع، قادر به تشخیص دقیق حروف در شرایط مختلف، نظیر نویز بالا و کیفیت پایین تصاویر، هستند

## 3-3-مروری بر مطالعات مرتبط:

مطالعات و کارهای نسبتاً کمی در زمینه پردازش تصویر و بینایی ماشین بر روی برچسب ارزش محصولات غذایی انجام گرفته است که همه این مطالعات بر روی برچسب‌هایی غیر از برچسبی فارسی پویش می‌شود بوده. بعضی از این محصولات به صورت نرم افزار اندرویدی[[65]](#footnote-65) بوده[9]، و برخی دیگر مقالات علمی[10-12].

این برنامه‌ها راه‌های مختلفی برای به دست آوردن ورودی کاربر از اطلاعات تغذیه‌ای دارند. یکی راه حل‌های این برنامه‌ها، استفاده از اطلاعاتی است که از قبل توسط مجموعه ای از کاربران جمع آوری شده است. مصرف‌کننده می‌توانند محصول خود را با نام جستجو کند تا آمار مربوطه را بیابد، اما داده های جمع آوری شده ممکن است کمبود داشته باشند.

اطلاعات، ممکن است برای هر محصول تکراری باشد و این اطلاعات بایکدیگر تطابق نداشته باشند که باعث گیج شدن می‌شود. همچنین اطلاعات ممکن است کاملاً نادرست باشد زیرا کاربران به صورت دستی آنها را وارد کرده و هیچ تعهدی نسبت به وارد کردن ‌آن‌ها ندارند. راه حل های دیگری که پیاده‌سازی شده، حول ساخت یک پایگاه داده از محصولات شناخته‌شده می‌باشد.

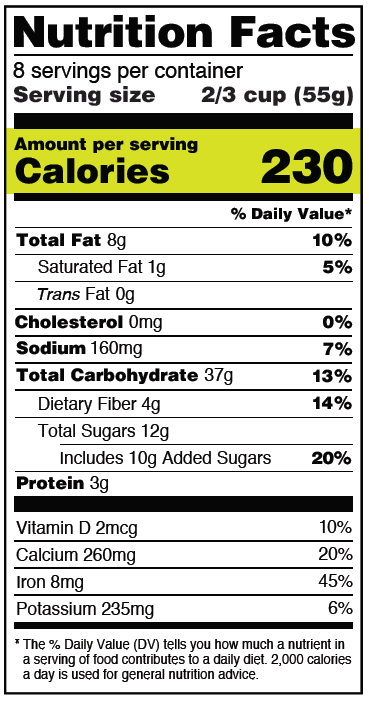
داده‌های تغذیه‌ای پایگاه‌داده محور، مستلزم جمع آوری تمام داده‌ها در یک مکان واحد است. داده‌های در حال تغییر باید دائما به روز شوند. این رویکرد به حجم زیادی از زیرساخت، همکاری و نگهداری نیاز دارد. اگر داده ها حفظ نشوند یا محصولات جدیدی به آن اضافه نشود، داده‌هایی که کاربر به دنبال آن است ممکن است وجود نداشته باشد یا قدیمی باشد.

روند کلی در مقالات مطالعه شده به صورت زیر است.

1. الگوریتم[[66]](#footnote-66)، یک عکس را به عنوان ورودی می‌گیرد.
2. برچسب ارزش مواد غذایی در عکس یافت شده و بریده می‌شود.
3. عکس به بخش‌های مختلفی تقسیم می‌شود که هر بخش نشانگر یک کلمه می‌باشد.
4. بخش‌های مختلف که عکس به آنها شکسته شد، خوانده شده تا عدد یا کلمه از آن استخراج شود.
5. خروجی به دست می‌آید و به صورت لیستی از اطلاعات نمایش داده می‌شود.

داده‌های استفاده شده در این مقالات، بسیار محدود تر از داده‌‌های موجود در الگوریتم معرفی شده در این مقاله می‌باشد. یکی از مقالات[10] به جای استفاده از OCR برای خواندن کلمات، از کلاس‌بند[[67]](#footnote-67) استفاده کرده به این دلیل که کلمات محدود بوده و نیاز به خواندن ندارند. ولی این کار باعث می‌شود که اگر یک کلاس‌بندی به اشتباه انجام شد، آن ماده مغذی به اشتباه مقدار بگیرد اما با استفاده از OCR، اگر مشکلی در خواندن پیش بیاید، یک یا دو حرف عوض شده و می‌توان از سایر حروف، کلمه را حدث زد. البته اگر تعداد عکس‌های مجموعه داده زیاد باشد، مشکل OCR تا حدودی رفع خواهد شد.

این تحقیقات عمدتاً برو روی برچسب محصولات آمریکایی انجام شده است که به شکل زیر است.



10- برچسب ارزش مواد غذایی آمریکایی

این برچسب شامل اطلاعات بیشتری نسبت به اطلاعات بر روی محصولات غذایی ایرانی دارد. این اطلاعات شامل، کالری هر سهم از محصول، تعداد وعده ها به ازای هر محصول، انواع چربی‌ها، کلسترول[[68]](#footnote-68)، سدیم[[69]](#footnote-69)، کربوهیدرات کل[[70]](#footnote-70)، پروتئین[[71]](#footnote-71)، ویتامین د[[72]](#footnote-72)، کلسیم[[73]](#footnote-73)، آهن[[74]](#footnote-74)و پتاسیم[[75]](#footnote-75)، می‌باشد. همچنین مقادیر این مواد مغذی هم به میزان وزنی و هم به درصد نوشته شده است.

یکی از تفاوت های اصلی برچسب‌های آمریکایی و ایرانی در رنگی بودن پس‌زمینه مواد مغذی می‌باشد. این رنگی بودن، کار خواندن را کمی دشوار تر از برچسب‌های آمریکایی کرده به این دلیل که کد رنگی RGB هر رنگ ممکن است با کد رنگی مشکی به سختی قابل تمایز بخشیدن باشد و مدل[[76]](#footnote-76) باید داده‌های زیادی را ببیند تا بتواند بین رنگ‌ها و کدهای رنگی تمایز قائل شود.

در برخی مطالعات انجام شده[12]، تمرکز بر روی خواندن برچسب با چولگی تا 30-45 درجه می‌باشد. همچنین قابلیت خواندن بارکد نیز با همین چولگی نیز میسر است. اما در این پروژه، تمرکز فقط بر روی خود برچسب بدون چولگی بیش از حد می‌باشد.

## 4-3- جمع بندی:

### ۱-۲-۲- تراز پلاک[[77]](#footnote-77):

اولین پروژه انجام شده در مدت زمان حضور کارآموزی، پروژه‌ی ترازسازی پلاک بود. در این پروژه که برای بهبود تصاویر در راستای تشخیص بهتر متن پلاک انجام شد، در ابتدا سعی شد تا توسط الگوریتم‌های سنتی از جمله خط‌واصل[[78]](#footnote-78) خطوط منحنی پلاک بدست‌آید و سپس توسط پرسپکتیو[[79]](#footnote-79) نمودن آن، پلاکی تراز شده و صاف حاصل گردد. به دلیل شرایط محیطی متغیر، این الگوریتم منجر به خروجی مناسبی نمی‌شد و برای همین پس از ۱ هفته تحقیق و پژوهش به سراغ یادگیری عمیق رفته و شبکه‌ای برای تشخیص ۴ گوشه پلاک پیاده‌سازی و آموزش داده شد.

### ۲-۲-۲- برنامه برچسب‌گذاری پلاک:

پس از تهیه و تست اولیه مدل ترازسازی پلاک، نیاز به جمع‌آوری مجموعه‌داده‌ی بیشتر، نمایان شد و طبیعتاً هر مجموعه‌ی داده برای وظایف بانظارت[[80]](#footnote-80) نیاز به برچسب[[81]](#footnote-81) دارد؛ برای همین، برنامه‌ای جهت برچسب‌گذاری دستی گوشه‌های پلاک نوشته شد تا بتوان با سرعت و دقت بالایی برچسب‌گذاری تصاویر رو انجام داد.

### ۳-۲-۲- رفع نویز و شفاف‌سازی پلاک:

در ادامه به سراغ یکی دیگر از مشکلات که کیفیت پایین بعضی از ورودی‌های مدل تشخیص متن بود رفته و سعی شد تا کیفیت عکس‌ها را به صورت در-لحظه[[82]](#footnote-82) بهبود دهیم. در ابتدا به سراغ الگوریتم‌های سنتی از جمله اعمال فیلتر گاووسی و جمع تصویر با لبه‌های آن، برای بهبود کیفیت و رفع تار و ماتی[[83]](#footnote-83) تصویر رفته اما به خروجی با تغییر قابل محسوسی دست نیافتیم. بنابراین به سراغ یادگیری عمیق رفته و شبکه‌ای پیچشی[[84]](#footnote-84) برای این کار طراحی گردید؛ به دلیل پیچیدگی بالای مسئله، شبکه‌های پیچشی توان کدنمودن[[85]](#footnote-85) مناسب تصویر را برای انجام چنین وظیفه‌ای[[86]](#footnote-86) نداشته و باید از شبکه‌هایی پیچیده‌تر برای کدنمودن تصویر استفاده می‌شد که به دلیل اینکه این کار به صورت در-لحظه امکان پذیر نبود، این پروژه در همین نقطه به پایان رسیده و به زمانی دیگر موکول شد.

### ۴-۲-۲- دنبال‌کننده پلاک:

در نهایت آخرین پروژه در دوران کارآموزی، پروژه‌ی اضافه نمودن دنبال‌کننده به خط اجرایی[[87]](#footnote-87) نرم‌افزار تشخیص پلاک بود که برای این کار از دنبال‌کننده‌ی مرکزگرا[[88]](#footnote-88) استفاده شد. در این پروژه مشکل تعداد بالای پلاک‌ها توسط دنبال‌کننده حل شد و طی الگوریتمی تنها یک تصویر برای ذخیره نمودن در پایگاه‌داده[[89]](#footnote-89) انتخاب شده و از تعداد بالای تصاویر و ذخیره‌سازی بیهوده‌ی تصاویر اجتناب گردید.

### ۵-۲-۲- انتقال برنامه‌های نوشته‌شده به محیط‌کاری دیپ‌استریم[[90]](#footnote-90):

در آخر دوره کارآموزی، تحقیقاتی پیرامون محیط‌کاری دیپ‌استریم صورت گرفت و سعی‌شد تا مدل‌های شرکت بر بستر این محیط پیاده‌سازی شوند تا سرعت و توزیع بار بهبود یابد. این بخش از پروژه همچنان در حال انجام می‌باشد و به اتمام نرسیده‌است.

## ۱-۳- مقدمه:

در این فصل به ارزیابی کلی دوره کارآموزی و کیفیت این دوره پرداخته می‌شود و سعی می‌شود تا مزایا و معایب این دوره در شرکت دیدپردازش هوپاد بیان شود تا دیگران بتوانند درباره‌ی این دوره با دیدی بازتر تصمیم‌گیری نمایند.

## ۲-۳- تجارب:

در این بخش به تجارب کسب شده در طی این دوره کارآموزی پرداخته می‌شود و سعی می‌شود تا دید بهتری به خواننده، درباره‌ی بازارکار و دانشگاه داده‌شود.

می‌توان گفت یکی از مهم‌ترین تجارب و آموخته‌های این دوره برای اینجانب، پی بردن به رابطه‌ی صنعت و دانشگاه و مهم‌تر از آن، کسب اعتمادبنفس بود. تا قبل از این دوره، رابطه‌ی صنعت با دانشگاه برای من به شفافی پس از آن نبود و اکنون با مشکلات صنعت تا حدودی آشنا شده‌ام. در طی این دوره، پی بردم که برخلاف تصورات قبلی خود، با توانایی‌های موجود نیز می‌توانم در صنعت فعالیت نمایم. یکی از تجربه‌های این دوره، کسب مهارت‌های نرم برای تعالم با سایر افراد تیم و نیز انجام کار تیمی به صورت جدی بود که به جد می‌توان گفت کار تیمی در صنعت فاصله‌ی زیادی با کار تیمی در دانشگاه دارد.

همچنین، می‌توان این مورد را پی برد که بیشتر رابطه‌ی صنعت و دانشگاه در ایران تنها در تبادل دانشجو محدود گشته و دانشگاه‌ها تلاش زیادی برای حمایت و هدایت آموزش‌ها و پژوهش‌ها به سمت صنعت و حل مشکلات نمی‌کنند. آموزش‌های دانشگاهی برای داشتن اطلاعات اولیه درباره‌ی موضوعات مختلف، مناسب هستند اما به تنهایی نمی‌توانند مهارت لازم در فرد را برای حل مشکلات صنعت ایجاد نمایند و خوب است دانشگاه با برگزاری دوره‌هایی عملی برای دانشجویان، هرچه بیشتر سعی در افزایش مهارت فنی دانش‌آموختگان خود نماید.

## ۳-۳- دست‌آوردهای فنی:

یکی از مهم‌ترین دلایلی که دوره کارآموزی گذرانده می‌شود، افزایش مهارت‌های فنی فرد در این دوره می‌باشد. در ادامه به شرحی از مهارت‌های کسب‌شده در این دوره پرداخته می‌شود.

در این دوره توانستم مهارت‌های خود را در موضوع پردازش‌تصویر، هوش‌مصنوعی و یادگیری‌عمیق افزایش دهم و شبکه‌هایی مختلفی برای وظایف مختلفی طراحی و آموزش دهم. در اولین پروژه این دوره، پس از عدم توانایی الگوریتم‌های سنتی در ترازنمودن پلاک‌ها، برای این کار به سراغ شبکه‌های عصبی رفته و شبکه‌ای برای این کار طراحی و آموزش داده شد و به نتایج بسیار خوبی دست یافته‌شد. در ادامه، برای افزایش کیفیت عکس‌ها به سراغ شبکه‌های آماده رفته و آن‌ها تست شدند که به دلیل زمان‌بر بودن آن‌ها و عدم تولید خروجی در لحظه، باز به سراغ طراحی شبکه توسط خودمان رفتیم؛ به دلیل پیچیدگی بالای این وظیفه، شبکه خروجی مناسبی تولید ننموده و این کار به آینده موکول شد.

یکی دیگر از دست‌آوردهای اینجانب در این دوره، نوشتن برنامه‌ای برای برچسب‌گذاری تصاویر بود که در طی این کار توانستم با رخدادها[[91]](#footnote-91) در کتابخانه‌ی opencv آشنا شوم. همچنین، در این دوره انواعی از الگوریتم‌های سنتی به کار گرفته‌شده و مهارت‌هایم در پردازش تصویر به مراتب بیشتر از قبل شد.

همان‌طور که پیش‌تر ذکرشد، در این دوره، مهارت کار با کتابخانه‌های tensorflow و pytorch را در خود افزایش دادم و اکنون راحت‌تر از قبل می‌توانم شبکه‌های عصبی را آموزش دهم، همچنین نحوه بهینه‌سازی[[92]](#footnote-92) شبکه‌های عصبی را آموختم و نحوه‌ی تبدیل مدل‌های pytorch به onnx را فراگرفتم.

یکی از مهم‌ترین مهارت‌های فنی کسب‌شده در این دوره برای من، کار با سیستم‌عامل لینوکس[[93]](#footnote-93) و محفظه‌های داکر[[94]](#footnote-94) بود که موجب شد، اطلاعات بیشتری درباره‌ی آن‌ها کسب نمایم.

یکی از جالب‌ترین آموخته‌های من در این دوره، توانایی حل مسئله بود که به دلیل سیاست‌های شرکت در من ایجاد شد! در مورد این موضوع در بخش‌های بعدی مفصلا توضیح داده ‌خواهد‌شد.

## ۴-۳- موفقیت‌ها و ناکامی‌ها:

در این بخش به بیان مشکلات و موفقبت‌های فنی در این دوره طی شده، پرداخته خواهد‌شد.

در این دوره، پروژه‌های ترازسازی پلاک، افزودن دنبال‌کننده مرکزگرا و نیز، چند منبعی نمودن کد آن‌ها با توجه به مهارت‌هایی که داشته و کسب نمودم، با موفقیت انجام شدند. اما، پروژه افرایش کیفیت عکس و کاهش نویز، به دلیل پیچیدگی بالای مسئله و عدم داشتن دانش و ومهارت‌های لازم برای طراحی شبکه‌های پیچیده‌تر و قوی‌تر، به نتیجه‌ی قابل استفاده‌ای منجر نشد. برای طراحی انجام چنین وظیفه‌ای لازم است تا شبکه توانایی بسیار بالایی در کدگذاری تصویر داشته باشد که این کار نیازمند استفاده از المان‌های پیچیده‌تر در شبکه است.

## ۵-۳- ارزیابی:

در این بخش به ارزیابی کلی از محل کارآموزی و توصیه‌های لازم به افرادی که قصد انتخاب این محل را برای گذراندن دوره کارآموزی خود دارند، پرداخته می‌شود.

اینجانب این محل را حتماُ به افرادی که قصد آموختن مطالب توسط خودشان و انجام پروژه‌های واقعی که در شرکت حتی استفاده می‌شوند و نه صرفاً مطالعه کاربرگ آموزشی و انجام تمارینی در حد دانشگاه، دارند پیشنهاد می‌نمایم. در این محل وظایف از کارهای خود شرکت بوده و فرد پروژه‌ای واقعی که در صنعت مورد نیاز است را انجام خواهد داد.

یکی از جالب‌ترین و بهترین نکات شرکت، وجود ارائه‌های آموزشی و معرفی تکنولوژی‌های مختلف توسط افراد در روزهای هفته می‌باشد که موجب می‌شود فرد با ایده‌های مختلف آشنا شده و سطح اطلاعاتی وسیعی درباره‌ی موضوعات مختلف کسب نمایند.

یکی دیگر از دلایل توصیه این شرکت، وجود سیستم ارزیابی و نمره‌دهی به فرد در طی یک ماه است که توسط این سیستم، فرد بر اساس پارامترهایی سنجیده می‌شود. این شرکت علاوه‌بر موارد گفته‌شده دارای محیطی شاد و دوستانه‌ است که موجب کاهش خستگی کار می‌شود.

# فصل چهارم

# گزارش‌های فنی

## ۱-۴- مقدمه:

در این فصل به شرح جزئیات پروژه‌های انجام شده پرداخته می‌شود و سعی می‌شود تا اطلاعات کاملی راجع‌به آن‌ها بیان شود. همچنین، اطلاعات درباره‌ی الگوریتم‌های به‌کار رفته، شرح داده‌می‌شوند.

## ۲-۴- تراز‌سازی پلاک:

به عنوان اولین پروژه، وظیفه ترازسازی عکس پلاک‌ها به اینجانب سپرده شد. در این پروژه، برنامه‌ای باید نوشته می‌شد تا عکسی از پلاک دریافت نموده و سپس آن‌را به صورت ترازشده برگرداند. در ابتدا، سعی شد تا این پروژه توسط الگوریتم‌های سنتی پردازش تصویر انجام شود. برای این کار، در ابتدا توسط الگوریتم کشف کناره[[95]](#footnote-95) لبه‌های پلاک کشف و مشخص شده و سپس از روی آن ۴ گوشه پلاک استخراج می‌شوند. سپس توسط محاسبه‌ی ماتریس هوموگرافی[[96]](#footnote-96) تصویر پلاک را به چهارگوشه[[97]](#footnote-97) تصویر نمادهی[[98]](#footnote-98) می‌دهیم. در ادامه، شرحی از نحوه کار الگوریتم کشف کناره خواهیم داد.

### ۱-۲-۴- کشف کناره‌ها:

برای کشف کناره‌ها از الگوریتم آماده‌ی کتابخانه‌ی اپن‌سی‌وی[[99]](#footnote-99) استفاده شد که عملکرد توسط آقای سوزوکی[[100]](#footnote-100) طراحی شده است. این الگوریتم به عنوان ورودی تصویر دودویی[[101]](#footnote-101) یا تصویر لبه‌یابی شده توسط الگوریتم همچون کنی[[102]](#footnote-102)، دریافت کرده و سپس کناره‌ها را به عنوان خروجی به صورت لیست برمی‌گرداند. این بردارها[[103]](#footnote-103) دارای مختصات شروع و پایان کناره هستند.

### ۲-۲-۴- الگوریتم پیدانمودن مستطیل:

در ادامه، توسط تابع آماده اپن‌سی‌وی از روی خروجی کناره‌ها، مستطیل و ۴ گوشه استخراج شد. برای داشتن خروجی بهتر، یک آستانه[[104]](#footnote-104) برای آن‌ها بر روی مساحت و داشتن حتمی ۴ گوشه قرار داده شد. این تابع به این صورت عمل می‌نماید که توسط یک فیلتر[[105]](#footnote-105)، عکس ورودی توسط عملگرهای ریخت‌شناسی[[106]](#footnote-106) فیلتر شده و تنها کناره‌هایی باقی می‌مانند که شکل مستطیلی دارند.

پس از طی تمامی این مراحل، خروجی مناسبی به دلیل وجود نویز و نور بدست نیامد. پس به سراغ تغییر الگوریتم مورد استفاده گرفتیم. در روش جدید، به جای پیدا نمودن کناره‌ها سعی شد تا نقاط کلیدی[[107]](#footnote-107) روی تصویر توسط مقایسه‌ی آن با یک تصویر قالب[[108]](#footnote-108) بدست آید. برای این کار الگوریتم‌های متعددی از جمله موارد زیر موجود می‌باشند:

* DoG
* Harris
* GFFT
* ORB

این الگوریتم‌ها همگی نقاط کلیدی تصویر را پیدا می‌نمایند. حال لازم است تا این نقاط با نقاط کلیدی قالب مقایسه شوند و سپس توسط این مقایسه تصویر دوباره با نمادهی تصویر، تراز شود. برای بدست آوردن توصیف‌گرهای نامتغیر محلی[[109]](#footnote-109) نیز الگوریتم‌های متعددی از جمله موارد زیر بدست می‌آیند:

* SIFT
* SURF

در آخر نیز برای نظیر کردن نقاط بدست آمده از الگوریتم RANSAC استفاده شد. این کار نیز به دلیل مواجه با شرایط نوری متغیر با شکست روبه‌رو شد و برای همین به سراغ یادگیری عمیق و شبکه‌های عصبی رفتیم.

### ۳-۲-۴- تراز‌سازی توسط شبکه‌های عصبی عمیق:

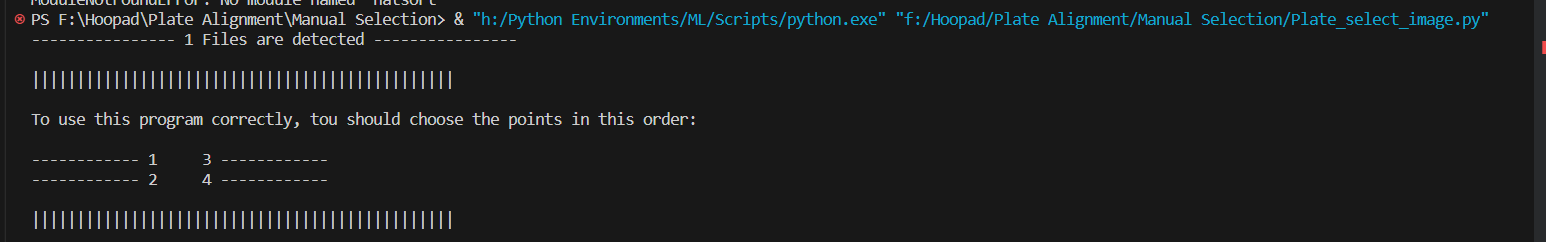
در ابتدا به دلیل سادگی این وظیفه، سعی شد تا تنها با قرار دادن ۳ لایه پیچشی و یک شبکه عصبی چندلایه[[110]](#footnote-110) این کار را حل کرد. معماری ابتدایی شبکه به صورت زیر می‌باشد:



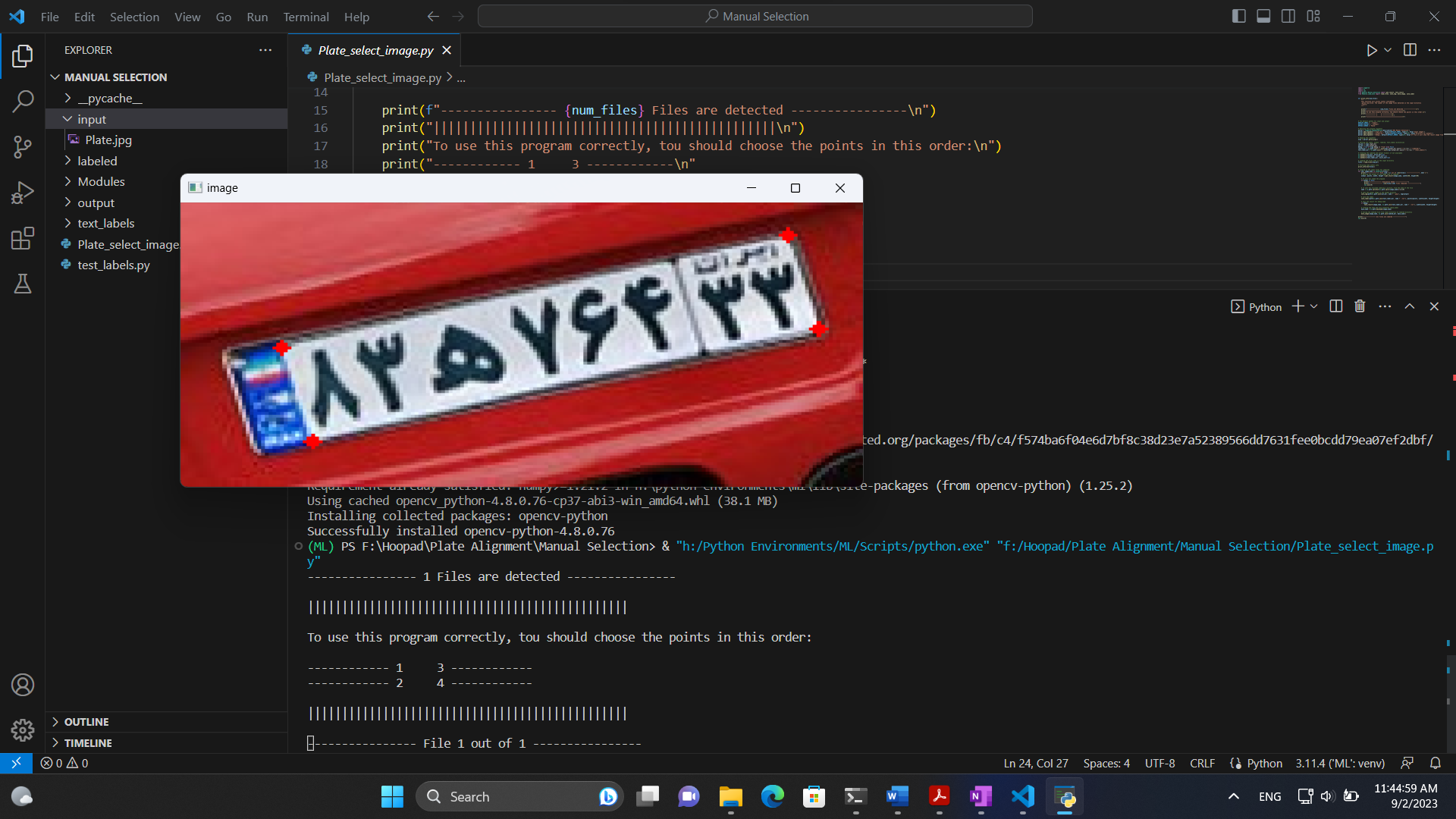
تصویر ۴-۱- شبکه ۳ لایه برای ترازسازی

برای عملکرد درست شبکه نیاز به داشتن حجم داده‌ی زیاد یک واقعیت انکارناپذیر است. برای همین، یک برنامه در راستای برچسب‌گذاری ۴ گوشه پلاک‌ها طراحی و توسعه داده‌شد تا بتوان مجموعه‌های داده را با سرعت بیشتری آماده نمود.

برای برچسب‌گذاری با چند چالش مواجه بودیم. اندازه متفاوت تصاویر، موجب یکسان نشده مقیاس گوشه‌های پلاک می‌شد که همین موجب عدم امکان واکشی مجدد ۴ گوشه پلاک از روی مختصات موجود می‌شد. برای حل این مشکل، به جای ذخیره‌سازی مختصات ۴ گوشه، مختصات بر طول و عرض تصویر تقسیم شدند تا تمامی پارامترها به عددی بین ۰ و ۱ تبدیل شوند و در نتیجه بتوان برای تمام تصاویر از یک مقیاس استفاده نمود. در ادامه تصاویری از این برنامه‌ را مشاهده می‌نمایید:



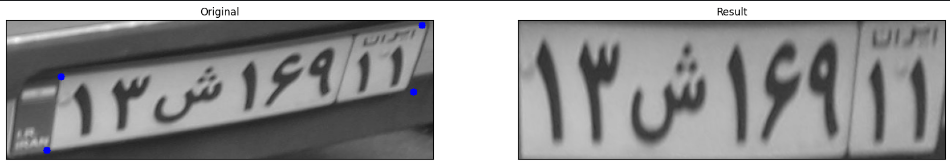
تصویر ۴-۲- راهنمای برنامه‌ی برچسب‌گذاری



تصویر ۴-۳- محیط برنامه‌ی برچسب‌گذاری

شاید سوال باشد که چرا در لایه آخر شبکه طراحی شده، ۸ نورون قرار داده‌شده است. علت آن است که ما ۴ نقطه خواهیم داشت که هر یک دارای طول و عرض می‌باشند. برای اینکه خروجی شبکه نیز حتماً بین ۰ و ۱ قرارگیرد از تابع فعال‌سازی[[111]](#footnote-111) سیگموید[[112]](#footnote-112) استفاده شد.

در نهایت، توسط این شبکه تواتسیم به خروجی‌های مناسبی در شرایط مختلف نوری دست یابیم که در ادامه یک مثال از عملکرد آن‌را مشاهده می‌نمایید:



تصویر ۴-۴- نمونه پلاک ترازشده

این مدل توانست به شاخص آردو[[113]](#footnote-113) پنجاه درصد دست یابد. در ادامه این شاخص و نحوه محاسبه‌ی آن شرح داده می‌شود.

### ۴-۲-۴- شاخص آردو:

این شاخص با نام ضریب اطمینان[[114]](#footnote-114) نیز شناخته می‌شود. این ضریب عددی بین ۰ و ۱ است که نشان می‌دهد یک مدل رگرسیون، چقدر اعداد را به خوبی تخمین می‌زند. به صورت کلی می‌توان چنین فرمولی را برای این شاخص بیان نمود:

در این رابطه، RSS بیان‌گر جمع مربعات باقی‌مانده[[115]](#footnote-115) است که از رابطه‌ی زیر بدست می‌آید:

که در آن fi برابر با پیش‌بینی مدل، yi برابر با خروجی واقعی در نقطه‌ی iام می‌باشد. همچنین TSS برابر با جمع کل مربعات می‌باشد که از رابطه‌ی زیر بدست می‌آید:

که در آن برابر با میانگین کل مقادیر است[13].

[1] General Department of Supervision and Evaluation of Food Products and F. a. D. O. cosmetics, "Instruction and guide for inserting nutritional color indicator." [Online]. Available: <https://fdo.skums.ac.ir/Dorsapax/Data/Sub_5/File/neshangar%20rangi.pdf>

[2] "Roboflow." <https://roboflow.com/> (accessed.

[3] "Label studio." <https://labelstud.io/> (accessed.

[4] "Saitak number convertor." <https://www.saitak.com/number> (accessed.

[5] "Kaggle." <https://www.kaggle.com/> (accessed.

[6] "EasyOCR repository." <https://github.com/JaidedAI/EasyOCR> (accessed.

[7] "Craft repository." <https://github.com/clovaai/CRAFT-pytorch> (accessed.

[8] "Deep text recognition benchmark repository." <https://github.com/clovaai/deep-text-recognition-benchmark> (accessed.

[9] "Scan Nutrition Labels to Track Foods Faster." <https://macrofactorapp.com/label-scanner-announcement/> (accessed.

[10] J. REIBRING, "Photo OCR for Nutrition Labels," 2017. [Online]. Available: <https://odr.chalmers.se/bitstream/20.500.12380/256649/1/256649.pdf>

[11] R. S. Nate Matsunaga, "Image processing for the extraction of nutritional

information from food labels," Santa Clara University, 2015. [Online]. Available: <https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1041&context=cseng_senior>

[12] T. Zaman, "Vision Based Extraction of Nutrition Information from Skewed Nutrition Labels," 2016. [Online]. Available: <https://digitalcommons.usu.edu/etd/4893>

[13] S. Turney. "Coefficient of Determination (R²) | Calculation & Interpretation." Scribbr. <https://www.scribbr.com/statistics/coefficient-of-determination/> (accessed.

1. Image processing [↑](#footnote-ref-1)
2. Computer vision [↑](#footnote-ref-2)
3. Labeling [↑](#footnote-ref-3)
4. Detector [↑](#footnote-ref-4)
5. Models [↑](#footnote-ref-5)
6. Image annotation [↑](#footnote-ref-6)
7. Bounding boxes [↑](#footnote-ref-7)
8. Segmentation masks [↑](#footnote-ref-8)
9. key points [↑](#footnote-ref-9)
10. XML [↑](#footnote-ref-10)
11. JSON [↑](#footnote-ref-11)
12. Platform [↑](#footnote-ref-12)
13. Roboflow [↑](#footnote-ref-13)
14. Label studio [↑](#footnote-ref-14)
15. Open source [↑](#footnote-ref-15)
16. API [↑](#footnote-ref-16)
17. Saitak [↑](#footnote-ref-17)
18. OpenCV [↑](#footnote-ref-18)
19. Intel [↑](#footnote-ref-19)
20. Linux [↑](#footnote-ref-20)
21. MacOS [↑](#footnote-ref-21)
22. PIL (Python Imaging Library) [↑](#footnote-ref-22)
23. GPJ [↑](#footnote-ref-23)
24. PNG [↑](#footnote-ref-24)
25. GIF [↑](#footnote-ref-25)
26. BMP [↑](#footnote-ref-26)
27. Pillow [↑](#footnote-ref-27)
28. NumPy [↑](#footnote-ref-28)
29. Ndarray [↑](#footnote-ref-29)
30. Data mining [↑](#footnote-ref-30)
31. RE [↑](#footnote-ref-31)
32. Regular Expressions [↑](#footnote-ref-32)
33. Format reshaper [↑](#footnote-ref-33)
34. Arabic reshaper [↑](#footnote-ref-34)
35. Bidi (Bidirectional) [↑](#footnote-ref-35)
36. RTL [↑](#footnote-ref-36)
37. Ultralytics [↑](#footnote-ref-37)
38. YOLO (You Only Look Once) [↑](#footnote-ref-38)
39. PyTorch [↑](#footnote-ref-39)
40. Facebook [↑](#footnote-ref-40)
41. Tensor [↑](#footnote-ref-41)
42. Graph [↑](#footnote-ref-42)
43. Numpy [↑](#footnote-ref-43)
44. Scikit-learn [↑](#footnote-ref-44)
45. CUDA (Compute Unified Device Architecture) [↑](#footnote-ref-45)
46. Nvidia [↑](#footnote-ref-46)
47. GPU [↑](#footnote-ref-47)
48. C [↑](#footnote-ref-48)
49. C++ [↑](#footnote-ref-49)
50. Tensorflow [↑](#footnote-ref-50)
51. Rendering [↑](#footnote-ref-51)
52. Kaggle [↑](#footnote-ref-52)
53. Notebook [↑](#footnote-ref-53)
54. Python [↑](#footnote-ref-54)
55. R [↑](#footnote-ref-55)
56. EasyOCR [↑](#footnote-ref-56)
57. Detection [↑](#footnote-ref-57)
58. Recognition [↑](#footnote-ref-58)
59. CRAFT (Character Region Awareness for Text detection) [↑](#footnote-ref-59)
60. Deep text recognition [↑](#footnote-ref-60)
61. Optical character recognition [↑](#footnote-ref-61)
62. USDA nutrition label [↑](#footnote-ref-62)
63. Convoloutional neural network [↑](#footnote-ref-63)
64. Recurrent neural network [↑](#footnote-ref-64)
65. Android [↑](#footnote-ref-65)
66. Algorithm [↑](#footnote-ref-66)
67. Classifier [↑](#footnote-ref-67)
68. Cholesterol [↑](#footnote-ref-68)
69. Sodiom [↑](#footnote-ref-69)
70. Total carbohydrate [↑](#footnote-ref-70)
71. Protein [↑](#footnote-ref-71)
72. Vitamin D [↑](#footnote-ref-72)
73. Calcium [↑](#footnote-ref-73)
74. Iron [↑](#footnote-ref-74)
75. Potassium [↑](#footnote-ref-75)
76. Model [↑](#footnote-ref-76)
77. Plate Alignment [↑](#footnote-ref-77)
78. Contour [↑](#footnote-ref-78)
79. Perspective [↑](#footnote-ref-79)
80. Supervised Task [↑](#footnote-ref-80)
81. Label [↑](#footnote-ref-81)
82. Real-Time [↑](#footnote-ref-82)
83. Blur [↑](#footnote-ref-83)
84. Convolutionary-Network [↑](#footnote-ref-84)
85. Encode [↑](#footnote-ref-85)
86. Task [↑](#footnote-ref-86)
87. Pipeline [↑](#footnote-ref-87)
88. Centroid Tracker [↑](#footnote-ref-88)
89. Database [↑](#footnote-ref-89)
90. DeepStream [↑](#footnote-ref-90)
91. Events [↑](#footnote-ref-91)
92. Optimization [↑](#footnote-ref-92)
93. Linux [↑](#footnote-ref-93)
94. Docker Containers [↑](#footnote-ref-94)
95. Contour [↑](#footnote-ref-95)
96. Homography [↑](#footnote-ref-96)
97. Frame [↑](#footnote-ref-97)
98. Perspective [↑](#footnote-ref-98)
99. OpenCV [↑](#footnote-ref-99)
100. Suzuki

     [suzuki\_et\_al.pdf (columbia.edu)](https://www.nevis.columbia.edu/~vgenty/public/suzuki_et_al.pdf) [↑](#footnote-ref-100)
101. Binary [↑](#footnote-ref-101)
102. Canny [↑](#footnote-ref-102)
103. Vectors [↑](#footnote-ref-103)
104. Threshold [↑](#footnote-ref-104)
105. Filter or kernel [↑](#footnote-ref-105)
106. Morphologically [↑](#footnote-ref-106)
107. Keypoints [↑](#footnote-ref-107)
108. Template [↑](#footnote-ref-108)
109. Local Invariant Descriptors [↑](#footnote-ref-109)
110. Multi-Layer Perceptron (MLP) [↑](#footnote-ref-110)
111. Activation Function [↑](#footnote-ref-111)
112. Sigmoid [↑](#footnote-ref-112)
113. R2 Score [↑](#footnote-ref-113)
114. Coefficient of Determination [↑](#footnote-ref-114)
115. Residual [↑](#footnote-ref-115)