دانشکده مهندسی کامپیوتر

پروژه پایانی کارشناسی

گزارش نهایی

تشخیص کاراکترهای دست‌نویس فارسی با استفاده از شبکه‌های عصبی

استاد راهنما : جناب آقای دکتر مزینی

نویسنده : محمدعلی فراهت

تاریخ انتشار : اسفند ۱۴۰۱

با تشکر و سپاس فراوان از استاد ارجمند پروژه کارشناسی جناب آقای دکتر ناصر مزینی برای وقت و راهنمایی‌هایی که در اختیار بنده قرار دادند.

تشخیص کاراکترهای دست‌نویس فارسی با استفاده از شبکه‌های عصبی

چکیده

تبدیل نوشته‌های دست نویس که در قالب تصویر ذخیره می‌شوند به متن فارسی تایپ شده همواره یکی از سخت‌ترین و پر چالش‌ترین کارهایی است که در در حوزه شبکه‌های عصبی و پردازش زبان‌های طبیعی وجود دارد. با توجه به پیشرفتی که در این زمینه در زبان‌های دیگر و به خصوص زبان انگلیسی به وجود آمده، این نیاز در زبان فارسی هم حس می‌شود. اما متاسفانه این مسئله در زبان فارسی پیچیدگی‌های بیشتری ایجاد می‌کند. زیرا حروف در زبان فارسی بر خلاف زبان انگلیسی به صورت جدا از هم نیستند و حتی ممکن است چندین شکل نوشتن داشته باشند، برای مثال حرف "ح" را ممکن است به چهار صورت "ح"، "حـ"، "ـح"و "ـحـ" در یک نوشته ببینیم.

در نتیجه این چالش‌ها باعث شده که در نتوانیم نتیجه خوبی در این زمینه برای زبان فارسی کسب کنیم. پس هرگونه پیشرفتی در این زمینه میتواند ما را یک قدم به رسیدن به این هدف نزدیک کند. اولین مرحله برای تشخیص متون فارسی، تشخیص کاراکترهای فارسی می‌باشد. در این تحقیق تلاش شده است تا با استفاده از شبکه‌های عصبی مختلف دقت این بازشناسی را بهبود ببخشیم.

فهرست مطالب

فصل ۱

مقدمه

۱/۱ مقدمه

با توجه به پیشرفت و توسعه تکنولوژی در دهه‌های اخیر به خصوص در زمینه رایانه‌ها و خدماتی که آن‌ها می‌توانند به بشر ارائه بدهند شاهد استفاده روز افزون از آن‌ها می‌باشیم. اخیرا این امر در حوزه هوش مصنوعی و یادگیری ماشینی خود را بیشتر به چشم ما نشان می‌دهد. آخرین تکنولوژی که اخیرا هم معرفی شده، سیستم ChatGPT می‌باشد که توجه زیادی را به خود جلب کرده و با استفاده از هوش مصنوعی سیستم سوال و جواب به کاربران ارائه می‌دهد.

امروزه بسیاری از نیازهای مردم وابستگی زیادی به موبایل‌ها و رایانه‌ها دارد. بسیاری از کلاس‌های مدارس و دانشگاه‌ها، جلسات شرکت‌ها و همچنین دیگر ارتباطات انسان‌ها به صورت مجازی و با استفاده از سیستم‌های رایانه‌ای اتفاق می‌افتد. به این دلیل، هرگونه اقدامی برای راحت‌تر کردن این امور برای انسان‌ها می‌تواند بسیار مورد استقبال قرار بگیرد.

یکی از این اقدام‌ها می‌تواند سیستمی باشد که بتواند متون دست‌نویس را به متون تایپی تبدیل کند. این سیستم‌های توسعه‌یافته با عنوان OCR[1] معرفی شده‌اند. این کار می‌تواند باعث شود بسیاری از کارهای ما سریع‌تر، دقیق‌تر و با هزینه کمتری انجام شود. این سیستم می‌تواند در بسیاری از شرکت‌ها مثل بانک‌ها استفاده شود و از تلف شدن وقت برای تایپ مجدد یک متن جلوگیری کند. یا مثلا در سیستم‌ها پلاک خوان پلیس که کاربردهای بسیار متنوعی دارد، میتواند استفاده شود.

استفاده از همچین سیستمی می‌تواند علاوه بر مزیت‌های ذکر شده در بالا، باعث پیشرفت‌های دیگری در حوزه‌های مختلف شود. بسیاری از اطلاعات مهم و کاربردی که در اینترنت موجود است به صورت متن تایپی نیست و با استفاده از رایانه نمی‌توان اطلاعات لازم را از آن‌ها استخراج کرد. اما با استفاده از همچنین سیستمی می‌توانیم حجم اطلاعات مفید و قابل استفاده را در اینترنت را بسیار بیشتر کنیم و آن‌ها برای یادگیری و تحلیل به ماشین‌ها بدهیم.

تمام این کاربردهایی که ذکر شد به شرطی قابل استفاده است که سیستم بتواند قابل اعتماد باشد، یعنی از عملکرد آن در شرایط مختلف مطمئن باشیم. در حال حاضر این اطمینان برای این سیستم‌ها در زبان فارسی وجود ندارد. در این پروژه قصد داریم تا گامی کوچک در راستای بهبود آن برداریم.

[1] Optical Character recognition

۱/۲ بیان مسئله تحقیق

به علت پیچیدگی بالای این سیستم‌ها، نیاز است که سیستم را به مراحل کوچکتری تقسیم کنیم و قدم به قدم جلو برویم. اولین قدم در این سیستم این است که بتوانیم حروف فارسی را با دقت خوبی تشخیص دهیم. تا در مراحل بعدی بتوانیم کلمات و سپس جملات دست‌نویس فارسی را تشخیص دهیم. پس ما در این تحقیق تمرکز خود را روی شبکه‌هایی قرار می‌دهیم که بتوانند حروف دست‌نویس فارسی را تشخیص دهند.

فصل ۲

مروری بر منابع

۲.۱ پیشینه تحقیق

این پروژه بر اساس و ادامه پژوهش تشخیص متون دست‌نویس فارسی با استفاده از شبکه‌های عصبی، نوشته سرکار خانم کاشانیان می‌باشد. در تحقیق ایشان به مباحث بیشتری در این زمینه پرداخته شده بود اما ما، همانطور که در بخش قبل ذکر شد، در اینجا روی تشخیص کاراکترهای فارسی تمرکز کردیم و مدل‌های بیشتر و متنوع‌تری را ارائه دادیم و از بقیه بخش‌ها صرف نظر کردیم تا در این بخش عمیق‌تر شویم.

از جمله مباحثی که در تحقیق قبلی مورد پژوهش قرار گرفته بود می‌توان به تشخیص کلمه‌های فارسی و کاراکترها و کلمات عربی اشاره کرد. دلیل انتخاب زبان عربی در پژوهش ایشان، وجود شباهت بسیار زیاد در حروف زبان عربی و فارسی می‌باشد.

در زمینه طبقه‌بندی تصاویر دست‌نویس بسیاری از پژوهشگران تمرکز خود را بر روی تشخیص ارقام گذاشته‌اند و تا کنون مقالات زیادی در این رابطه منتشر شده. خوشبختانه در زمینه ارقام، کاراکترهای زبان فارسی و انگلیسی از لحاظ فاصله‌گذاری تفاوت زیادی ندارند و می‌توان از ایده‌ها و پیشرفت‌های محققان خارجی استفاده کرد. اما همانطور که اشاره شد، ما در این تحقیق فقط به تشخیص کاراکترها می‌پردازیم.

مقاله [1] که برای تشخیص حروف در زبان عربی در سال ۲۰۲۲ منتشر شده، کمک شایانی به من در رابطه با این تحقیق کرد. نحوه استفاده از دیتاست، پیش‌پردازش‌ها[2] و همچنین انتخاب مدل در این مقاله به زبانی ساده انجام شده بود که بسیار برای من مفید واقع شد. در این مقاله با استفاده از مدل پیشنهادی خود، موفق شدند که دقت تشخیص حروف را در دیتاست AHCD[3] از ۹۴.۹ درصد به ۹۶.۷۸ درصد افزایش دهند.

در تحقیق خانم کاشانیان، دقت تشخیص حروف فارسی در بهترین شبکه‌ای که استفاده کرده بودند در آموزش ۹۹.۴ درصد و در اعتبارسنجی ۹۴.۸ درصد بود. در دیتای تست که دست‌خط خودشان بود، ۷۸ درصد دقت کسب کرده بودند. در فصل‌های جلوتر میزان پیشرفت حاصل شده را نسبت به این تحقیق مقایسه خواهیم کرد.

در ادامه به بررسی مفاهیم پایه‌ای که برای فهم این تحقیق مورد نیاز است می‌پردازیم.

[2] Pre-processing

[3] Arabic handwritten characters dataset

۲.۲ شبکه‌های عصبی

شبکه‌های عصبی مصنوعی[۴] به زبان ساده‌تر شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت بیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. ایده اصلی این گونه شبکه‌ها تا حدودی الهام‌گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده‌ها و اطلاعات به منظور یادگیری و ایجاد دانش می‌باشد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است. فلسفه اصلی شبکه عصبی مصنوعی، مدل کردن ویژگی‌های پردازشی مغز انسان برای تقریب زدن روش‌های معمول محاسباتی با روش پردازش زیستی است. به بیان دیگر، شبکه عصبی مصنوعی روشی است که دانش ارتباط بین چند مجموعه داده را از طریق آموزش فراگرفته و برای استفاده در موارد مشابه ذخیره می‌کند.

یک شبکه عصبی مصنوعی، از سه لایه ورودی، خروجی و پردازش تشکیل می‌شود. هر لایه شامل گروهی از سلول‌های عصبی (نورون) است که عموماً با کلیه نورون‌های لایه‌های دیگر در ارتباط هستند، مگر این که کاربر ارتباط بین نورون‌ها را محدود کند؛ ولی نورون‌های هر لایه با سایر نورون‌های همان لایه، ارتباطی ندارند.

نورون کوچک‌ترین واحد پردازشگر اطلاعات است که اساس عملکرد شبکه‌های عصبی را تشکیل می‌دهد. یک شبکه عصبی مجموعه‌ای از نورون‌هاست که با قرار گرفتن در لایه‌های مختلف، معماری خاصی را بر مبنای ارتباطات بین نورون‌ها در لایه‌های مختلف تشکیل می‌دهند. نورون می‌تواند یک تابع ریاضی غیرخطی باشد، در نتیجه یک شبکه عصبی که از اجتماع این نورون‌ها تشکیل می‌شود، نیز می‌تواند یک سامانه کاملاً پیچیده و غیرخطی باشد. در شبکه عصبی هر نورون به‌طور مستقل عمل می‌کند و رفتار کلی شبکه، برآیند رفتار نورون‌های متعدد است. به عبارت دیگر، نورون‌ها در یک روند همکاری، یکدیگر را تصحیح می‌کنند.

شبکه عصبی کانولوشن[۵] نوع خاصی از شبکه عصبی با چندین لایه است که داده‌هایی را که آرایش شبکه‌ای دارند، پردازش کرده و سپس ویژگی‌های مهم آن‌ها را استخراج می‌کند. یک مزیت بزرگ استفاده از CNNها این است که نیازی به انجام پیش‌پردازش زیادی روی تصاویر نیست. یک تفاوت بزرگ بین CNN و شبکه عصبی معمولی این است که CNNها برای مدیریت ریاضیات پشت‌صحنه، از کانولوشن استفاده می‌کنند. حداقل در یک لایه از CNN، به‌جای ضرب ماتریس از کانولوشن استفاده می‌شود. کانولوشن‌ها تا دو تابع را می‌گیرند و یک تابع را برمی‌گردانند. این کار می‌تواند به صورت موازی انجام شود و در وقت صرفه‌جویی فراوانی صورت می‌گیرد.

[4] Artificial Neural Networks (ANN)
[5] Convolutional Neural Network (CNN)

۲.۳ معرفی دادگان

با بررسی‌ها و تحقیقاتی که انجام شد، برای آموزش مدل‌ها در این تحقیق از مجموعه داده "پایگاه داده جامع دستنویس برای تشخیص برون خط دستخط فارسی"[۶] استفاده شد.

این مجموعه داده در قسمت حروف فارسی دارای ۵۰۰ نویسنده بومی فارسی می‌باشد که نصف آن‌ها مرد و نصف دیگر زن هستند. تمام حروف به صورت فایل تصویری و خاکستری[۷] ذخیره شده‌اند و در پوشه‌های مربوط به کلاس خود قرار گرفته‌اند. با تغییراتی که من در کلاس‌ها دادم، در پایان ما دارای ۳۴ کلاس برای طبقه‌بندی هستیم که این ۳۴ کلاس برای هر حرف فارسی می‌باشند (حروف هـ و آ هم اضافه شده‌اند).

تعداد کل تصاویر ما ۱۷۰۰۰ عدد است و ما از ۸۰ درصد آن‌ها (۱۳۶۰۰ عدد) برای آموزش مدل ها و همچنین از ۲۰ درصد باقی‌مانده (۳۴۰۰ عدد) برای اعتبارسنجی[۸] مدل در هر ایپاک استفاده کردیم. در ضمن برای تست کردن مدل‌ها، برای اینکه به واقعیت نزدیک باشد، خودم و ۴ نفر دیگر، هرکدام یکبار تمام حروف را نوشتیم و با استفاده از آن‌ها دیتای تست را که شامل ۱۷۰ تصویر می‌باشد تولید کردیم.

برای آموزش مدل‌ها از افزایش داده‌ها[۹] استفاده می‌شد تا مدل فقط روی داده‌های خاص آموزش نبیند و در هر ایپاک آموزش ظاهر عکس‌ها کمی تغییر پیدا می‌کرد.

در زیر تعدادی از تصاویر اصلی را مشاهده می‌کنید:

تصویر ۱: تعدادی از داده‌های آموزش

[6] https://users.encs.concordia.ca/~j_sadri/Persian_Handwritten_Database.htm
[7] Grayscale
[8] Validation
[9] Data augmentation

فصل ۳

روش تحقیق

۳.۱ شبکه‌های عصبی استفاده شده

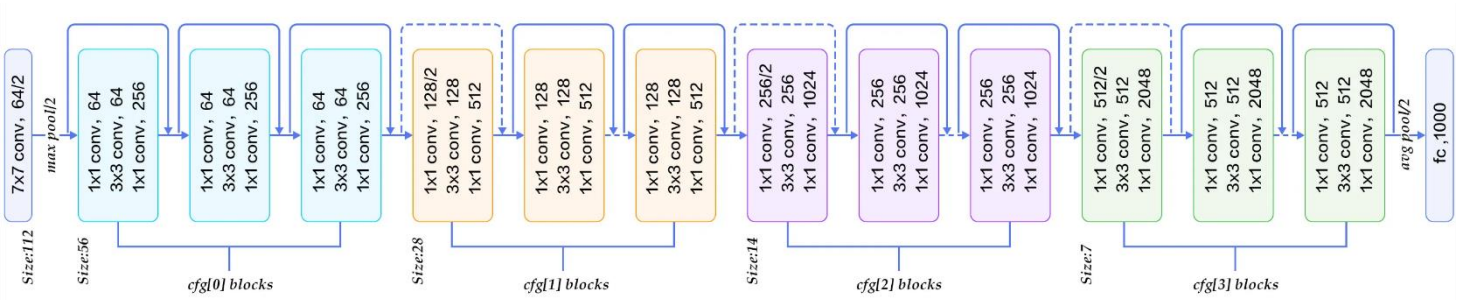
در این تحقیق هفت شبکه عصبی مختلف بررسی شد که بعضی از آن‌ها معروف بودند و بعضی از آن‌ها طراحی خودمان بود. برخی از این مدل‌ها پارامترهای بیشتری داشتند و حجم ذخیره شده آن‌ها بیشتر بود و طبیعتا زمان بیشتری برای آموزش نیاز داشتند. اما برخی هم کوچک‌تر بودند و سریع‌تر آموزش می‌دیدند.

برای آموزش این شبکه‌ها نیاز به پردازنده گرافیکی قدرتمندی بودیم که با استفاده از گوگل کولب[10] توانستیم این نیاز را برآورده کنیم. اما مشکلاتی مانند تمام شدن ظرفیت پردازنده گرافیکی آن و مشکلاتی نظیر اتصال اینترنت همواره باعث می‌شد که به فکر چاره باشیم. از این رو مدل‌ها در هر ایپاک ذخیره می‌شدند تا در صورت قطع شدن زمان اجرا، وزن‌های مدل ما از دست نروند.

در ادامه تمام مدل‌های بررسی شده و معماری آن‌ها را خواهیم دید.

۳.۱.۱ مدل ResNet50

این مدل که یک شبکه عمیق بسیار معروف است در سال ۲۰۱۵ معرفی شد و تا به آن سال جزو عمیق‌ترین شبکه‌های شناخته شده بود. این شبکه دارای ۱۵۲ لایه می‌باشد و به همین دلیل عملکرد بهتری روی تصاویر دارد. اما چالش طراحی آن در قسمت بهینه‌سازی است که در آن سال الگوریتم جدیدی برای این کار ارائه شد. معماری این شبکه را در تصویر زیر مشاهده می‌کنید.

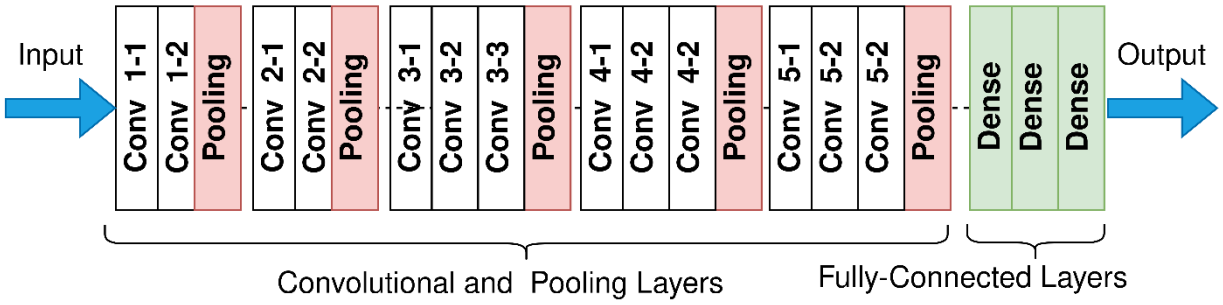


تصویر ۲: معماری شبکه ResNet50

[10] Google Colab

۳.۱.۲ شبکه VGG16

این شبکه یک سال قبل از شبکه ResNet و در سال ۲۰۱۴ معرفی شد. عملکرد آن در طبقه‌بندی تصاویر روی دیتاست ImageNet کمتر از شبکه ResNet می‌باشد اما به مراتب سبک‌تر است و عمق کمتری هم دارد. معماری ارائه شده برای این شبکه را در زیر می‌بینیم.



تصویر ۳: معماری شبکه VGG16

۳.۱.۳ شبکه CNN کوچک

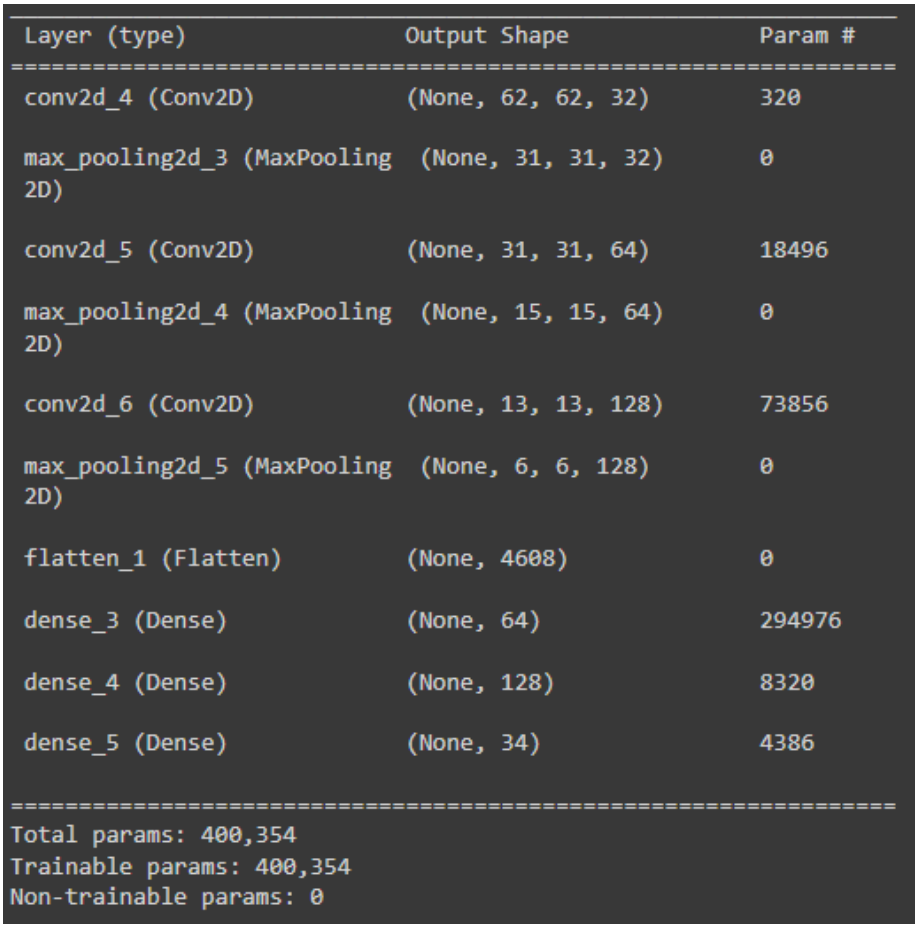
این مدل بسیار کوچک بوده و دارای ۳ لایه کانولوشنی و یک لایه dense برای طبقه بندی تصاویر به ۳۴ کلاس می‌باشد. معماری آن به صورت زیر است.

```
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d_17 (Conv2D)           (None, 60, 60, 128)       3328

 conv2d_18 (Conv2D)           (None, 58, 58, 64)        73792

 conv2d_19 (Conv2D)           (None, 56, 56, 32)        18464

 flatten_5 (Flatten)          (None, 100352)            0

 dense_6 (Dense)              (None, 34)                3412002

=================================================================
Total params: 3,507,586
Trainable params: 3,507,586
Non-trainable params: 0
```

تصویر ۴: معماری شبکه CNN کوچک

۳.۱.۴ شبکه CNN متوسط

این شبکه هم مانند شبکه قبلی طراحی ساده‌ای دارد اما با تعداد لایه‌های بیشتری ساخته شده. همچنین با وجود تعداد لایه بیشتر، می‌بینیم که به علت استفاده از pooling، تعداد پارامترهای آن بسیار کمتر شده. معماری آن را در تصویر زیر مشاهده می‌کنیم.



Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 62, 62, 32)	320
max_pooling2d_3 (MaxPooling 2D)	(None, 31, 31, 32)	0
conv2d_5 (Conv2D)	(None, 31, 31, 64)	18496
max_pooling2d_4 (MaxPooling 2D)	(None, 15, 15, 64)	0
conv2d_6 (Conv2D)	(None, 13, 13, 128)	73856
max_pooling2d_5 (MaxPooling 2D)	(None, 6, 6, 128)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_3 (Dense)	(None, 64)	294976
dense_4 (Dense)	(None, 128)	8320
dense_5 (Dense)	(None, 34)	4386

Total params: 400,354
Trainable params: 400,354
Non-trainable params: 0

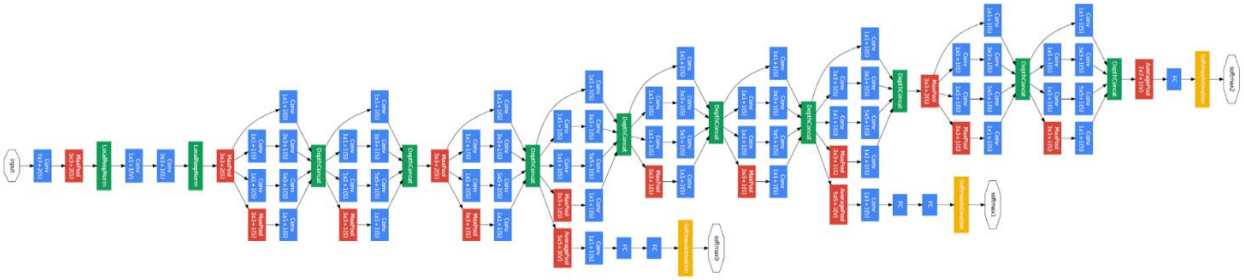
تصویر ۵: معماری شبکه CNN متوسط

۳.۱.۵ شبکه ResNet50 Pre-trained

این شبکه دقیقا همان شبکه ResNet اول می‌باشد اما با این تفاوت که وزن‌ها در آن به صورت تصادفی شروع نشدند و من از وزن‌های آماده آن برای دیتاست ImageNet استفاده کردم و دوباره مدل را آموزش دادم. با این کار احتمالا زمان کمتری برای آموزش مجدد و رسیدن به دقت بالاتر نیاز می‌باشد.

۳.۱.۶ شبکه InceptionV3

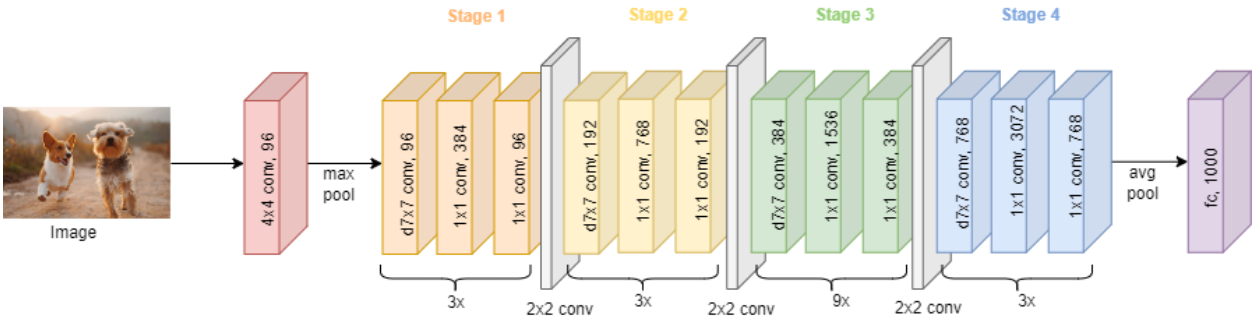
این شبکه بعد از GoogleNet و ورژن سوم آن در سال ۲۰۱۵ ارائه شد. این شبکه عمیق توانست دقت ۷۸ درصدی روی دیتاست ImageNet را به دست آورد. این شبکه در مجموع ۴۲ لایه دارد که برخی از آن‌ها موازی هم هستند و بعد از چند مرحله یکی می‌شوند (ایده آن از فیلم Inception گرفته شد). معماری آن را در زیر مشاهده میکنید.



شکل ۶: معماری شبکه InceptionV3

۳.۱.۷ شبکه ConvNeXtXLarge

این شبکه یکی از سنگین‌ترین و بزرگترین شبکه‌های موجود است. تعداد پارامترهای آن حدود ۳۴۸ میلیون است. آموزش آن هم به شدت زمان گیر است. این مدل در سال ۲۰۲۰ رونمایی شد و جزو جدیدترین مدل‌های معروف است. دقت این مدل بر روی دیتاست ImageNet برابر با ۸۶ درصد است که درصد بسیار خوبی است. معماری آن را در تصویر زیر می‌توان مشاهده کرد.



تصویر ۷: معماری شبکه ConvNeXtXLarge

فصل ۴

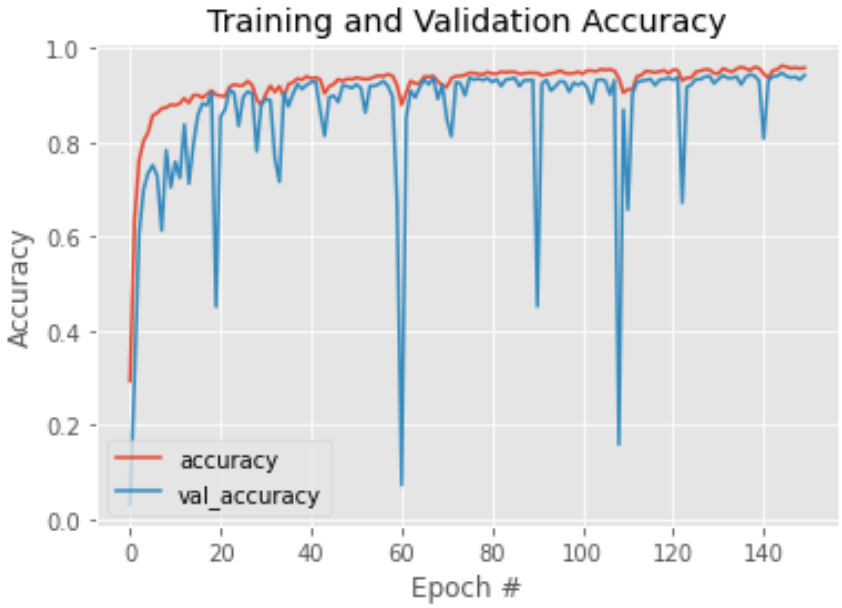
نتایج و تفسیر آنها

۴.۱ ارزیابی و مقایسه شبکه‌ها

آموزش شبکه‌ها با توجه به ورودی که برای هر شبکه تعریف کرده بودیم در تعداد ایپاک و batch size های مختلف تست کردیم و ساعت‌ها زمان برای آموزش مدل‌ها صرف کردیم. در زیر به طور مفصل آمار هرکدام از شبکه‌ها را شرح می‌دهیم.

۴.۱.۱ شبکه ResNet50

این شبکه به علت داشتن تعداد پارامترهای زیاد (۲۴ میلیون پارامتر) به زمان زیادی برای آموزش نیاز داشت. برای آموزش این مدل از تابع ضرر[11] categorical crossentropy استفاده کردم و همچنین optimizer مورد استفاده در این مدل و تمام مدل‌های دیگر هم Adam می‌باشد. این مدل بعد از آموزش در ۱۵۰ ایپاک و در حدود ۳ ساعت آموزش به دقت ۹۶ درصد در آموزش و ۹۴ درصد در اعتبارسنجی رسید. نمودار زیر گویای این اطلاعات است.

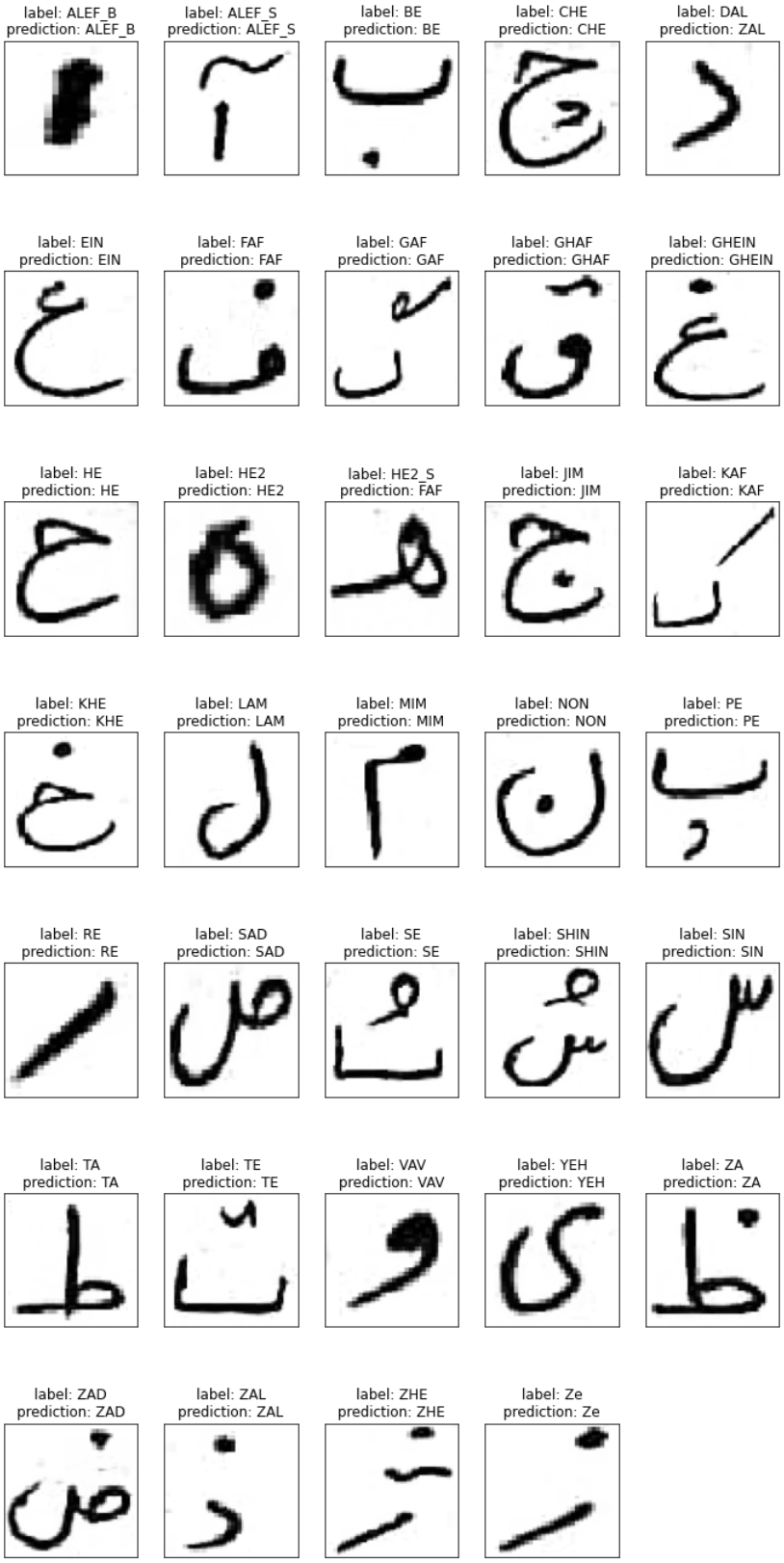


تصویر ۸: نمودار دقت در شبکه ResNet50

حجم نهایی مدل ذخیره شده حدود ۲۷۰ مگابایت بود.

این مدل روی دیتای تست دست‌نویس که توسط نویسنده این تحقیق و چهار نفر دیگر نوشته شده بود دقت بسیار خوب ۸۶.۴ درصد را داشت. در زیر نمونه‌ای از پیش‌بینی‌های این مدل را روی دیتای تست می‌بینید.

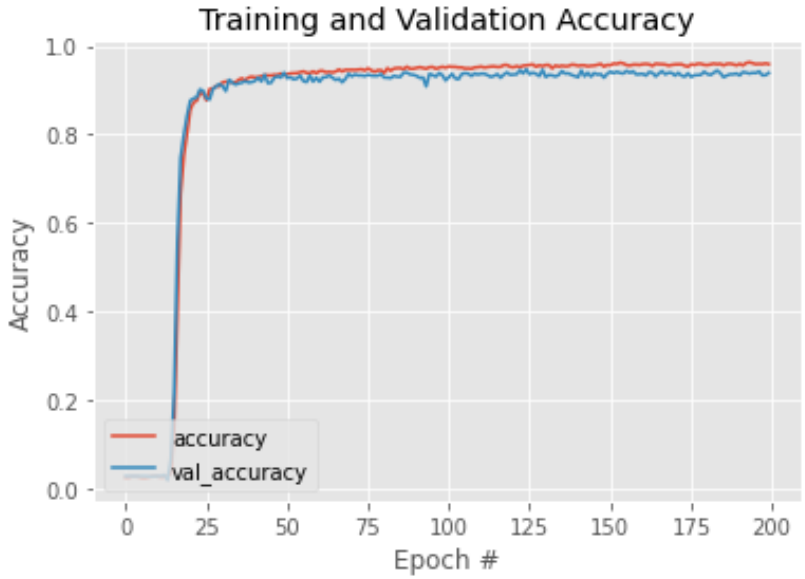
[11] Loss function



تصویر ۹: نمونه ای از پیشبینی مدل ResNet50 روی دیتای تست

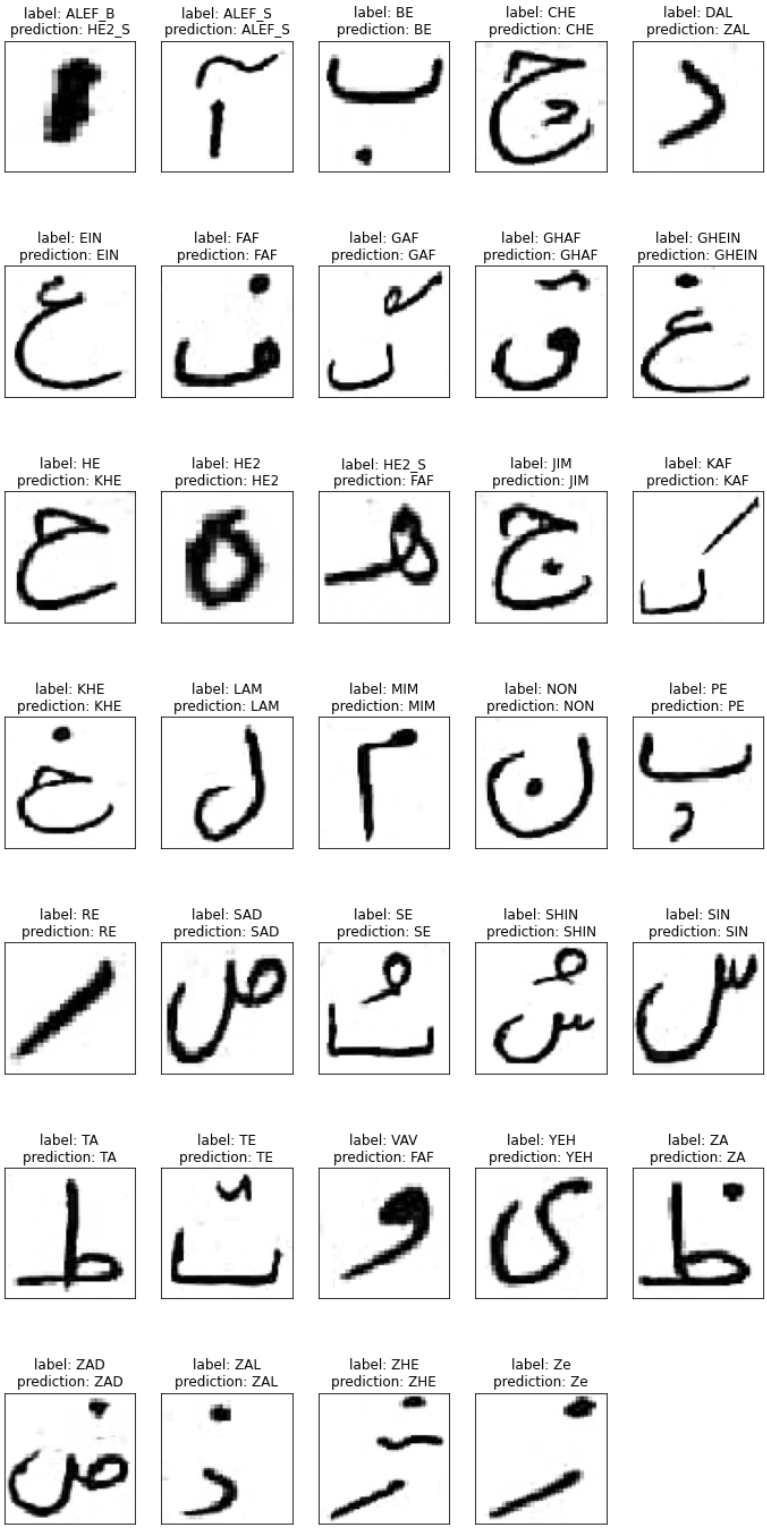
۴.۱.۲ شبکه VGG16

این شبکه دارای ۱۲ میلیون پارامتر بود و در ۲۰۰ ایپاک آموزش دید که این کار حدود ۴ ساعت زمان برد. دقت آن روی داده‌های آموزش به ۹۶ درصد و در داده‌های اعتبارسنجی به ۹۴.۷ درصد رسید. همچنین این مدل بر خلاف شبکه قبلی (ResNet50) نمودار یکنواخت‌تری داشت و مراحل آموزش به خوبی جلو میرفت. در زیر می‌توان این نمودار را مشاهده کرد.



تصویر ۱۰: نمودار دقت در شبکه VGG16

حجم نهایی مدل ۵۶ مگابایت شد که با توجه به تعداد پارامترهای آن منطقی است.
همچنین دقت مدل روی دیتا تست ۸۰.۵ درصد بود که از مدل ResNet کمتر می‌باشد. تعدادی از نمونه‌های پیشبینی این مدل را در زیر میتوانید مشاهده کنید.

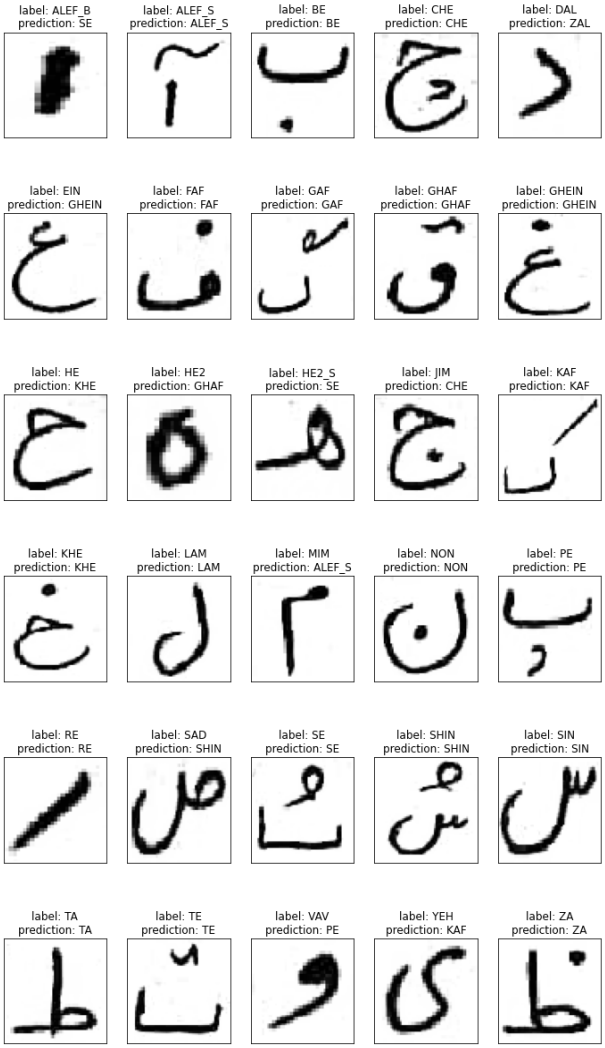


تصویر ۱۱: نمونه‌ای از پیشبینی شبکه VGG16 بر روی دیتای تست

۴.۱.۳ شبکه CNN کوچک

این شبکه که دارای ۳.۵ میلیون پارامتر بود در ۲۴۰ ایپاک حدود ۳ ساعت آموزش دید. در این دوره آموزش به دقت ۸۵ درصد در داده‌های آموزش و دقت ۸۴ درصد در داده‌های اعتبارسنجی رسید. متاسفانه به علت قطع شدن سیستم نتوانستیم آموزش را تا ۳۰۰ ایپاک که هدف ما بود انجام دهیم و همچنین نمودار دقت آموزش از دست رفت.

حجم این مدل ۴۰ مگابایت شد. همچنین دقت دیتا تست هم اصلا خوب نبود و ۴۸ درصد دقت داشت. میتوانیم این برداشت را بکنیم که این مدل به دلیل عمق کم و پارامترهای کم نتوانست به خوبی feature های تصاویر را پیدا کند و آموزش ببیند. تعدادی از پیشبینی‌های مدل را در زیر میتوانیم ببینیم و با مدل‌های قبلی مقایسه کنیم.



تصویر ۱۲: نمونه ای از پیشبینی مدل CNN کوچک بر روی دیتای تست

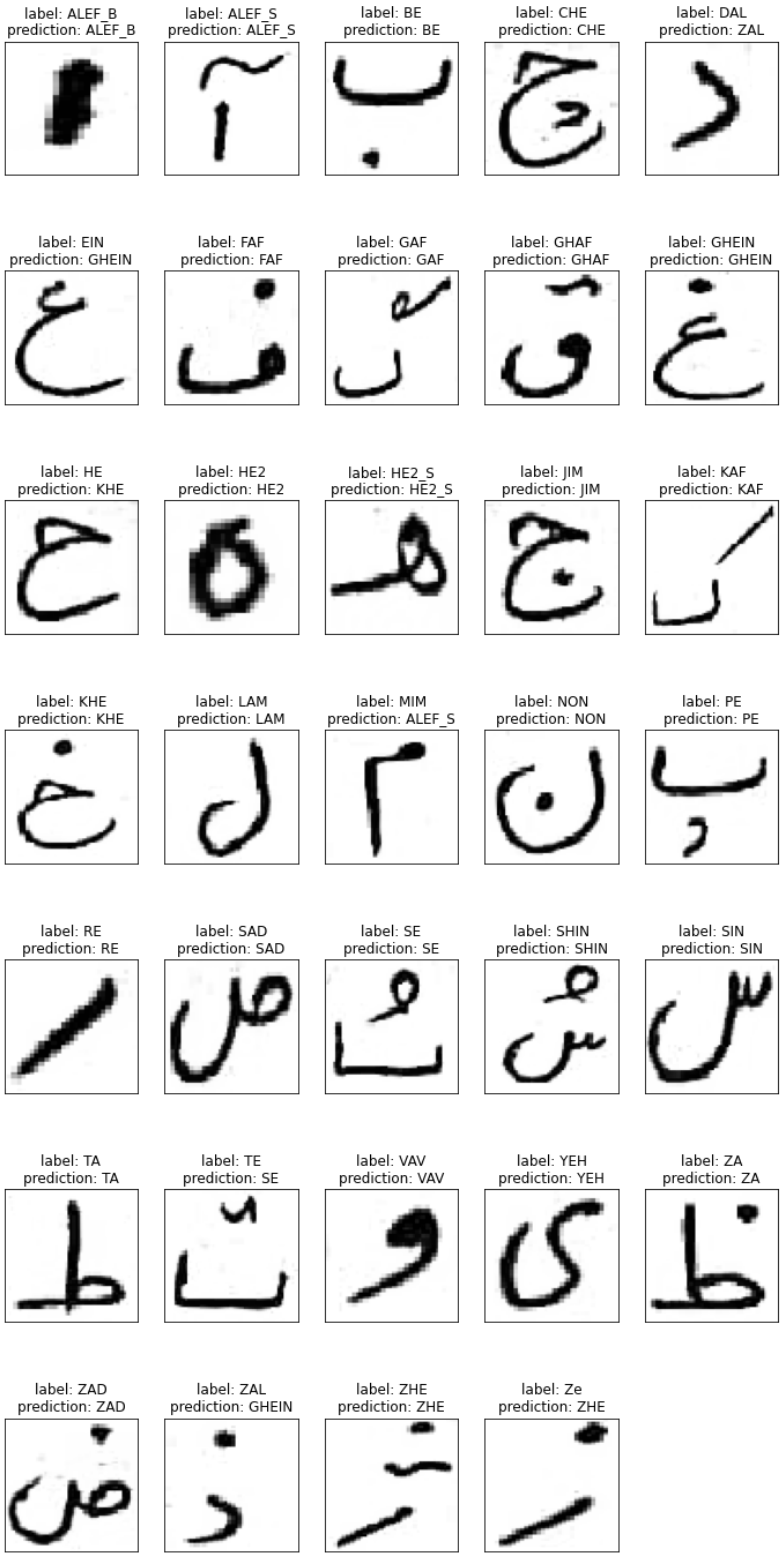
۴.۱.۴ شبکه CNN متوسط

این شبکه همانطور که در بخش قبل هم اشاره شد دارای ۴۰۰ هزار پارامتر می‌باشد. این مدل در ۳۰۰ ایپاک و حدودا ۴ ساعت آموزش دید تا به دقت ۹۲ درصد در داده‌های آموزش و ۹۱.۸ درصد در داده‌های اعتبارسنجی برسد. نمودار دقت آموزش این مدل را در این ۳۰۰ ایپاک در زیر می‌بینیم.

تصویر ۱۳: نمودار دقت شبکه CNN متوسط

این مدل به دلیل تعداد پارامتر کم، حجم کمی هم دارد و کمتر از ۲ مگابایت فضا اشغال کرده است. دقت این مدل در دیتا تست نزدیک به ۶۹ درصد است که با اینکه از همه بالاتر نیست، اما با توجه به تعداد پارامتر و مدت زمان آموزش، دقت نسبتا خوبی کسب کرده.

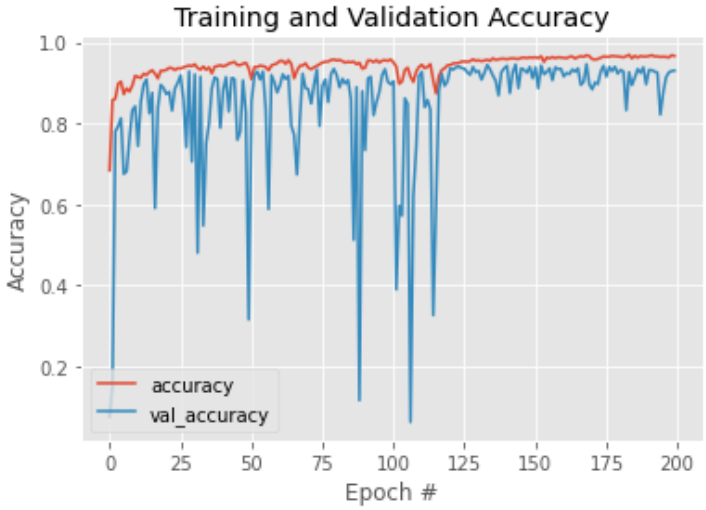
پیشبینی مدل برای برخی از دیتاهای تست را در تصویر زیر می‌توانید مشاهده کنید.



تصویر ۱۴: پیشبینی شبکه CNN متوسط بر روی دیتا تست

۴.۱.۵ شبکه ResNet Pre-trained

این شبکه کاملا شبیه به مدل ResNet می‌باشد. اما چون میخواهیم از وزن‌های آماده دیتاست ImageNet استفاده کنیم، باید تصاویر را رنگی داشته باشیم. برای همین ورودی را در ۳ کانال به مدل میدهیم و دیگر آن را grayscale نمی‌کنیم. این مدل را در ۲۰۰ ایپاک و مدت زمان بیشتر از ۴ ساعت آموزش می‌دهیم. در طول آموزش به دقت آموزش ۹۶.۶ درصد و به دقت اعتبارسنجی ۹۳ درصد می‌رسیم. نتیجه‌گیری که من انجام دادم این بود که علت کاهش دقت ما تصاویر رنگی بود، و نتیجه با تصاویر grayscale بسیار بهتر بودند. نمودار دقت در آموزش این شبکه را در زیر می‌بینیم.



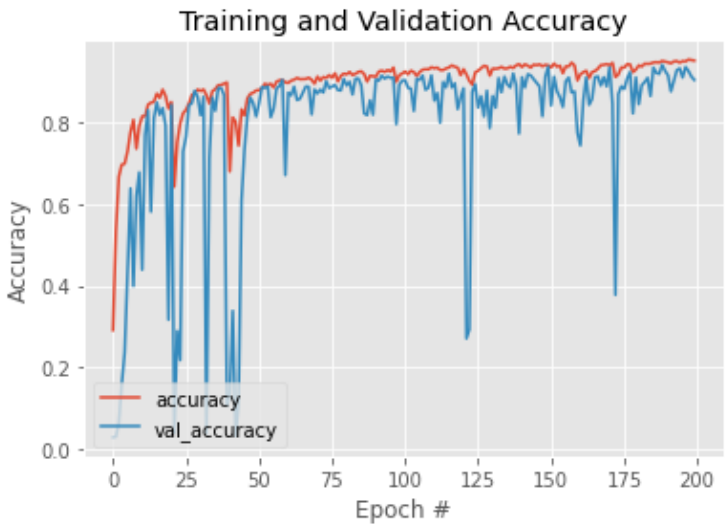
تصویر ۱۵: نمودار دقت در مدل ResNet Pre-trained

این مدل همانطور که انتظار میرفت، در تست هم اصلا خوب عمل نکرد و درحالی که همین مدل آموزش دیده شده روی تصاویر grayscale ، دقت بالای ۸۶ درصد را داشت، این مدل دقت ۶۵ درصد را روی دیتا تست دارد. نمونه‌ای از پیشبینی‌های این مدل را برای دیتا تست در زیر می‌بینیم.

تصویر ۱۶: نمونه‌ای از پیشبینی مدل ResNet Pre-trained برای دیتای تست

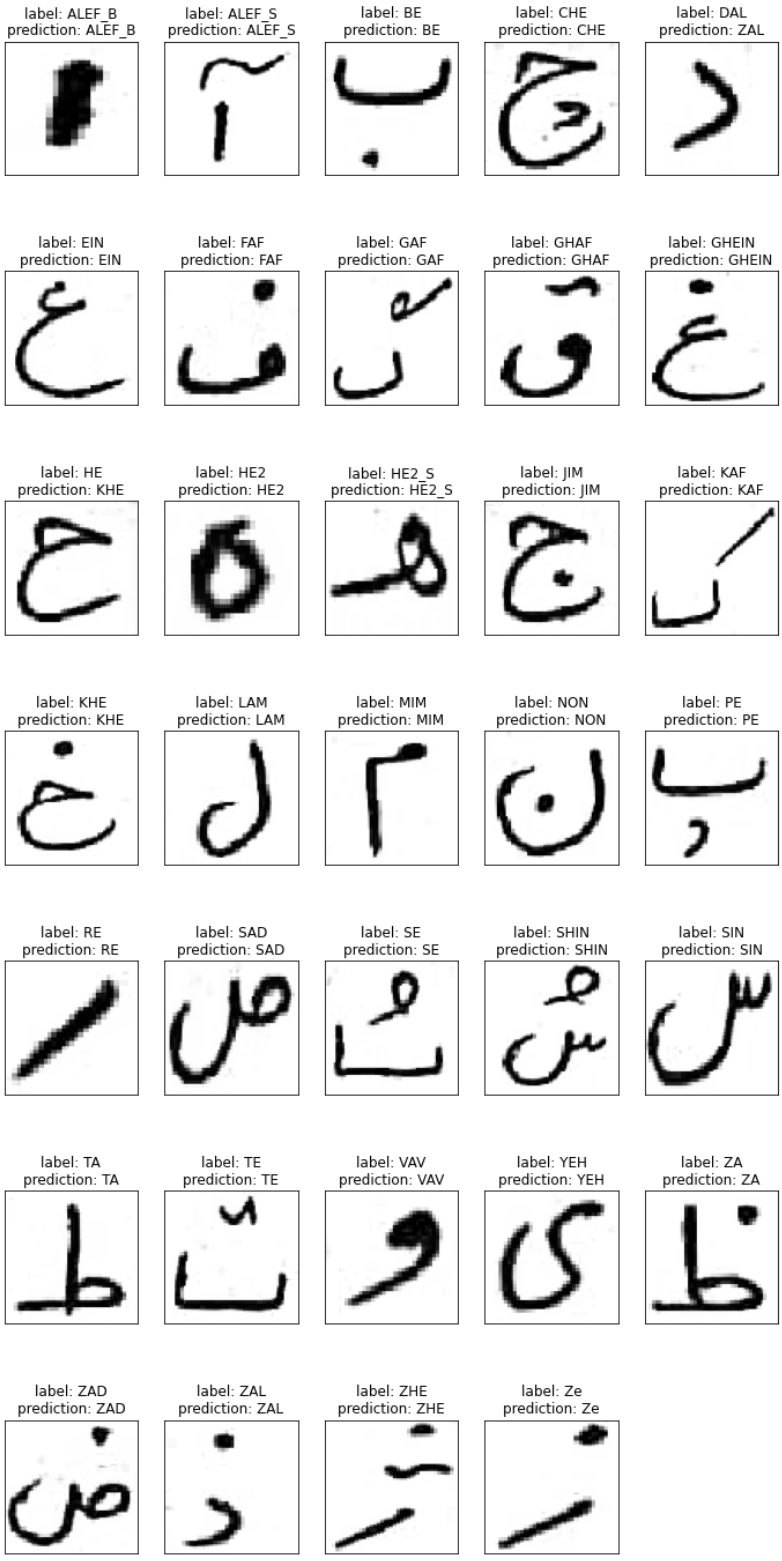
۴.۱.۶ شبکه InceptionV3

این شبکه که دارای ۲۲ میلیون پارامتر بود، در ۲۰۰ ایپاک و در مدت ۴ ساعت آموزش دید. نتیجه آن این شد که توانست به دقت ۹۵.۱ درصد در آموزش و دقت ۹۴ درصد در اعتبارسنجی برسد. ابعاد ورودی در این شبکه حداقل ۷۵ در ۷۵ بود که من هم همین مقدار را در نظر گرفتم. نمودار دقت را در این مدل میتوانیم در زیر ببینیم.



تصویر ۱۷: نمودار دقت شبکه InceptionV3 در طول آموزش

این مدل با دقت ۷۸ درصد در دیتا تست تا به اینجا توانست بعد از ResNet50 در رتبه دوم بایستد. در زیر نمونه‌ای از پیشبینی‌های این مدل را برای دیتای تست می‌بینیم.

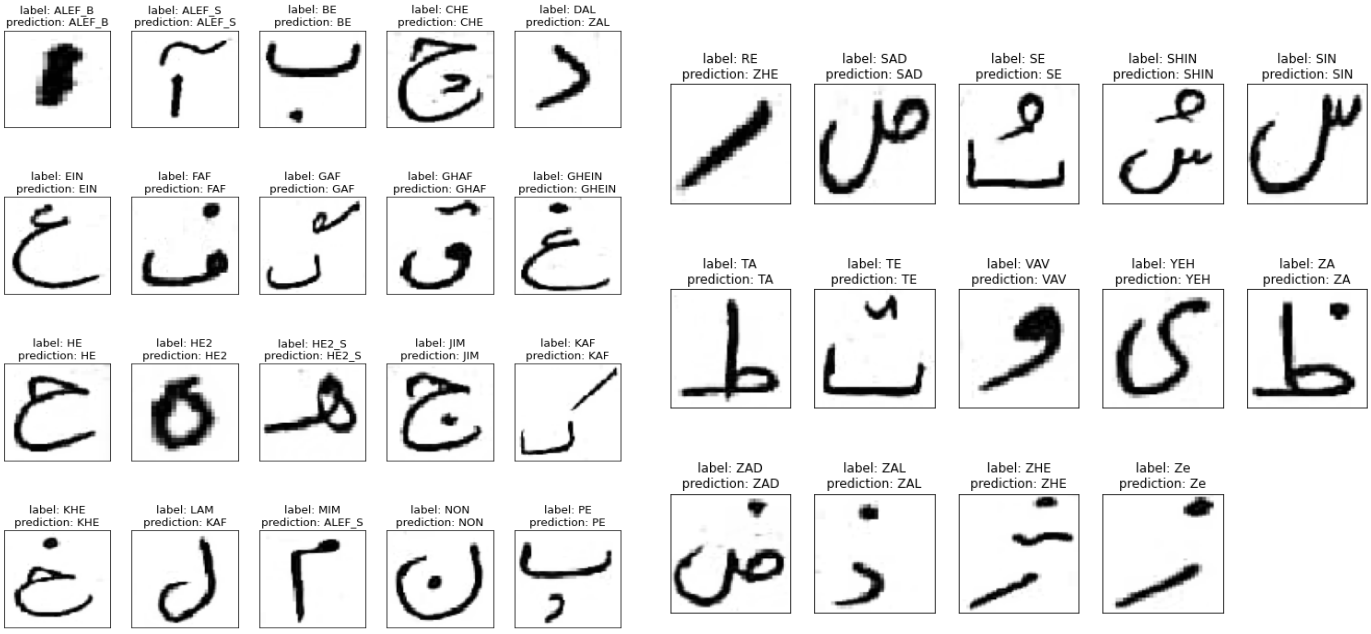


تصویر ۱۸: نمونه‌ای از پیشبینی‌های مدل InceptionV3 بر روی دیتای تست

۴.۱.۷ شبکه ConvNeXtXLarge

این شبکه که بزرگترین شبکه ما بود و با ۳۴۸ میلیون پارامتر حجمی نزدیک به ۴ گیگابایت را اشغال می‌کرد، توانست با ۹ ساعت آموزش در ۱۳۱ ایپاک به دقت ۹۳ درصد در دیتا آموزش و ۹۲ درصد در دیتا اعتبارسنجی برسد. که قطعا با مدت زمان آموزش بیشتر میتوانست دقت‌های بالاتری داشته باشد. متاسفانه به دلیل زمان بالای آموزش و قطع شدن سیستم، نمودار دقت آن از دست رفت.

این مدل با دقت بسیار خوب ۸۳.۵ درصد در دیتای تست توانست بعد از ResNet دوم شود و جای مدل InceptionV3 را بگیرد. در زیر میتوانیم تعدادی از پیشبینی‌های این مدل برای دیتای تست را ببینیم.



تصویر ۱۹: پیشبینی مدل ConvNeXtXLarge از نمونه‌هایی از دیتای تست

فصل ۵

جمع‌بندی و پیشنهادات

۵.۱ جمع‌بندی

در این تحقیق قصد این را داشتیم که اندکی به مدل ایده‌آلی نزدیک شویم که بتواند بدون خطا حروف فارسی را تشخیص دهد. طبق پژوهشی که انجام شد، از بین مدل‌های انتخاب شده، مدل ResNet50 بهترین عملکرد را بین بقیه داشت. حتی عملکرد آن نسبت به مدل بسیار سنگین ConvNeXtXLarge هم بهتر بود و این در صورتی است که ما محدودیت زمان و سخت افزار لازم برای اجرای آموزش‌ها را داشتیم.

اگر بخواهیم نتیجه این تحقیق را با نتیجه تحقیق قبلی مقایسه کنیم، میتوانیم بگوییم که دقت بهترین مدل ما، یعنی ResNet50 نسبت به بهترین مدل تحقیق قبلی یعنی مدل VGG، در دیتای تست توانست بیشتر از ۸ درصد بهبود داشته باشد.

دیگر نتایج قابل مقایسه را در جدول زیر می‌توان مشاهده کرد:

تحقیق	مدل	دقت آموزش	دقت اعتبارسنجی	دقت تست
قبلی	VGG	99.4	94.8	78.1
جدید	ResNet50	95.7	94.2	86.5

از جدول بالا می‌توان این نتیجه را گرفت که مدل استفاده شده تا حدودی Overfit شده و دقت بالا روی دیتای آموزش داشته اما روی داده‌های تست و اعتبار سنجی به مراتب ضعیف‌تر عمل کرده.

۵.۲ پیشنهادات

برای بهبود این پروژه چند پیشنهاد وجود دارد. بهتر بود که قبل از دادن ورودی تست ها به مدل، با استفاده از پردازش تصویر کمی کیفت حرف را بهبود ببخشیم تا شبیه به دیتایی شود که مدل از روی آن آموزش دیده. دیگر پیشنهاد من این است که لایه‌های آخر مدل ResNet استفاده شده را اندکی تغییر دهیم تا برای کار ما مناسب تر باشد.

از دیگر کارهایی که امکان انجام آن وجود داشت، استفاده از داده‌های بیشتر بود. برای حروف عربی مجموعه داده‌های بسیار زیادی وجود دارد و از بسیاری از آن‌ها میتوان به عنوان دیتای فارسی استفاده کرد چون شباهت زیادی دارند.

پیشنهاد آخر هم این است که با تغییر مدل پیشنهاد شده در مقاله [1] و با استفاده از دادگان فارسی آن را آموزش دهیم و احتمالا نتیجه خوبی خواهیم گرفت و در عین حال مدل بسیار بسیار ساده‌تر خواهد بود و امکان استفاده آن در دستگاه‌های دیگر مانند تلفن همراه هوشمند هم وجود دارد.

مراجع

[1] Ullah Z, Jamjoom M. 2022. An intelligent approach for Arabic handwritten letter recognition using convolutional neural network. PeerJ Computer Science 8:e995 https://doi.org/10.7717/peerj-cs.995

[2] Sadri, Javad, Mohammad Reza Yeganehzad, and Javad Saghi "A novel comprehensive database for offline Persian handwriting recognition." Pattern Recognition 60 (2016): 378-393.

تشخیص متون دست‌نویس فارسی با استفاده از شبکه‌های عصبی، نوشته سرکار خانم کاشانیان

https://keras.io/api/applications/

https://www.analyticsvidhya.com/blog/2020/08/image-augmentation-on-the-fly-using-keras-imagedatagenerator/

https://www.kaggle.com/code/suniliitb96/tutorial-keras-transfer-learning-with-resnet50

https://users.encs.concordia.ca/~j_sadri/PersianDatabase.htm