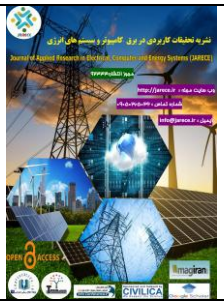
Vol. 2, No. 8, Jan-Mar 2025, (13-24)
Journal of Applied Research in Electrical, Computer and Energy Systems (JARECE)
نشریه تحقیقات کاربردی در برق ، کامپیوتر و سیستم های انرژی
دوره ۲، شماره ۸، زمستان ۱۴۰۳ (۲۴-۱۳)
journal homepage: https://jarece.ir/



شناسایی نویسه‌های دست‌نویس فارسی با بهره‌گیری از یک مدل ترکیبی عمیق مبتنی بر معماری‌های ResNet وEfficientNet

محمدمتین محمدی پگا[۱] ، امید طاهری[۲]

[۱] دانشجو، گروه کامپیوتر، دانشگاه ملی مهارت، تهران، ایران matinmmp1381@gmail.com
[۲] مدرس، گروه کامپیوتر، دانشگاه ملی مهارت، تهران، ایران omid.taheri6080@gmail.com

چکیده

بازشناسی نوری نویسه[۱] برای زبان فارسی به دلیل پیچیدگی‌های ساختاری خط، همواره یکی از حوزه‌های چالش‌برانگیز در هوش مصنوعی بوده است. در سال‌های اخیر، مدل‌های یادگیری عمیق[۲]، به‌ویژه شبکه‌های عصبی کانولوشنی[۳]، پیشرفت‌های چشمگیری در این زمینه داشته‌اند. این پژوهش، یک سامانه جامع برای تشخیص حروف دست‌نویس فارسی با استفاده از یک معماری ترکیبی و مدرن یادگیری عمیق ارائه می‌دهد. در این راستا، یک مجموعه‌داده بزرگ شامل 124 هزار تصویر از 120 کلاس مختلف حروف فارسی، با ترکیب دو مجموعه داده استاندارد ایجاد گردید. سپس، تصاویر با استفاده از یک الگوریتم پیش‌پردازش سفارشی[۴]، نرمال‌سازی شدند. ما در این پژوهش، یک معماری ترکیبی را بر پایه دو شبکه شناخته‌شده EfficientNet و ResNet-34 و با استفاده از تکنیک یادگیری انتقال[۵] بنا کرده‌ایم. برای افزایش مقاومت مدل در برابر بیش‌برازش[۶]، از تکنیک‌هایی همچون افزایش داده[۷] و توقف زودهنگام[۸] استفاده شد. نتایج ارزیابی نشان می‌دهد که مدل پیشنهادی توانسته است به دقت بیشینه 95.06% بر روی داده‌های اعتبارسنجی دست یابد. این عملکرد بالا، کارایی بالای رویکردهای نوین یادگیری عمیق را در حل مسئله پیچیده بازشناسی حروف فارسی اثبات می‌کند.

واژه‌های کلیدی: بازشناسی نوری نویسه‌ها، یادگیری عمیق، شبکه‌های عصبی کانولوشنی، یادگیری انتقال، معماری ترکیبی، حروف فارسی.

[1] Optical Character Recognition (OCR)
[2] Deep Learning
[3] Convolutional Neural Networks (CNN)
[4] Pre-processing
[5] Transfer Learning
[6] Overfitting
[7] Data Augmentation
[8] Early Stopping

۱ .مقدمه

زبان فارسی، با ویژگی‌های منحصربه‌فرد خود، چالش‌های خاصی را پیشروی سامانه‌های بازشناسی نوری نویسه (OCR) قرار می‌دهد. اتصال حروف به یکدیگر (حالت پیوسته)، تغییر شکل یک حرف بر اساس موقعیت آن در کلمه (آغاز، میانه، پایان و تنها)، وجود نقاط متعدد و حساس به مکان (مانند «ب»، «پ»، «ت» و «ث») و همچنین تنوع بالای سبک‌های دستخط، همگی باعث افزایش پیچیدگی مسئله تشخیص می‌شوند. این پیچیدگی‌ها باعث شده است که توسعه یک سامانه OCR دقیق و قوی برای زبان فارسی، همچنان یک حوزه تحقیقاتی فعال باقی بماند.

به‌طورکلی، رویکردهای تشخیص دست‌نویس به دودسته اصلی تقسیم می‌شوند: روش‌های آنلاین و روش‌های آفلاین. روش‌های آنلاین مبتنی بر تحلیل سیگنال‌های حرکتی هستند که در حین نوشتن، اطلاعات سنسورهایی مانند شتاب‌سنج را تحلیل می‌کنند [1]. در مقابل، روش‌های آفلاین که رویکرد اصلی این پژوهش است، پس از اتمام نگارش، تصویر نهایی را پردازش می‌نمایند. این رویکرد به دلیل کاربرد گسترده‌تر در اسناد موجود و عدم نیاز به تجهیزات خاص در زمان نگارش، از اهمیت بیشتری برخوردار است.

در دهه‌های گذشته، روش‌های کلاسیک یادگیری ماشین[9] بر پایه استخراج دستی ویژگی[10] از تصاویر استوار بودند. در این روش‌ها، متخصصان باید ویژگی‌های معناداری را از تصویر استخراج کرده و سپس آن‌ها را به یک طبقه‌بند ساده مانند درخت تصمیم[11] یا پرسپترون چندلایه[12] می‌دادند [2, 3]. عملکرد این روش‌ها به شدت به کیفیت ویژگی‌های استخراج‌شده وابسته بود و فرایند مهندسی ویژگی، خود امری زمان‌بر و پیچیده محسوب می‌شد.

با ظهور یادگیری عمیق، پارادایم جدیدی در حوزه بینایی ماشین[13] و OCR شکل گرفت. شبکه‌های عصبی کانولوشنی (CNN) قادرند به‌صورت خودکار و سلسله‌مراتبی، ویژگی‌های موردنیاز را مستقیماً از پیکسل‌های خام تصویر یاد بگیرند و نیاز به مهندسی ویژگی دستی را تا حد زیادی از بین ببرند. موفقیت چشمگیر معماری‌هایی مانند AlexNet [4] و پس از آن VGG [5] و ResNet [6]، راه را برای دستیابی به دقت‌های بی‌سابقه در وظایف مختلف بینایی ماشین هموار کرد. پژوهش‌های اخیر در حوزه OCR فارسی نیز به طور فزاینده‌ای به سمت استفاده از این معماری‌های عمیق حرکت کرده‌اند . [7]

هدف اصلی این پژوهش، طراحی و پیاده‌سازی یک سامانه OCR کارآمد برای حروف دست‌نویس فارسی با بهره‌گیری از یک معماری ترکیبی جدید است. در این مقاله، ما یک فرایند کامل را ارائه می‌دهیم که از مرحله ایجاد و پیش‌پردازش یک مجموعه‌داده بزرگ و جامع آغاز شده و به آموزش یک مدل ترکیبی که از دو معماری قدرتمند EfficientNet و ResNet-34 به‌عنوان استخراج‌کننده ویژگی استفاده می‌کند، ختم می‌شود. این رویکرد نوآورانه باهدف ترکیب نقاط قوت هر دو معماری برای دستیابی به‌دقت و کارایی بالاتر طراحی شده است.

[9] Machine Learning
[10] Manual Feature Extraction
[11] Decision Tree
[12] Multilayer perceptron
[13] Computer Vision

در ادامه مقاله، ابتدا در بخش دوم به بررسی مبانی نظری و کارهای پیشین در این حوزه می‌پردازیم. در بخش سوم، روش تحقیق شامل جزئیات مجموعه‌داده، مراحل پیش‌پردازش، معماری مدل پیشنهادی و فرایند آموزش به‌تفصیل شرح داده می‌شود. بخش چهارم به ارائه و تحلیل یافته‌های تجربی اختصاص دارد و در نهایت، بخش پنجم به نتیجه‌گیری و ارائه پیشنهاداتی برای کارهای آتی می‌پردازد.

۲ .مبانی تحقیق و پیشینه پژوهش

حوزه بازشناسی حروف دست‌نویس فارسی دارای پیشینه تحقیقاتی قابل‌توجهی است که می‌توان سیر تکامل آن را از روش‌های کلاسیک به سمت رویکردهای مدرن یادگیری عمیق مشاهده کرد.

۲.۱ .رویکردهای کلاسیک

رویکردهای اولیه عمدتاً بر پایه استخراج ویژگی‌های دستی و استفاده از طبقه‌بندهای آماری و یادگیری ماشین کلاسیک استوار بودند. برای مثال، مدحتی و همکاران [2] در پژوهش خود، سیستمی برای تشخیص حروف با استفاده از درخت تصمیم باینری ارائه دادند. آن‌ها ابتدا مجموعه‌ای از ویژگی‌های متنوع شامل فاصله، تبدیل فوریه کسینوسی[۱۴] و ویژگی‌های آماری را از تصاویر حروف استخراج کردند. این رویکرد، نمونه‌ای بارز از فرایند کلاسیک OCR است که در آن، موفقیت سیستم به شدت به توانایی محقق در طراحی و انتخاب ویژگی‌های متمایزکننده بستگی دارد. با پیشرفت یادگیری ماشین، محققان به سمت استفاده از شبکه‌های عصبی، حتی در قالب‌های ساده‌تر، حرکت کردند. توپچی و ابوالقاسم‌پور [3] در مقاله‌ای، از یک شبکه عصبی پرسپترون چندلایه[۱۵] برای تشخیص حروف فارسی بهره بردند. آن‌ها نیز در ابتدا با استفاده از روش ثابت‌های گشتاور[۱۶]، ویژگی‌هایی را از تصاویر استخراج کردند. این کار یک گام تکاملی نسبت به روش‌های پیشین بود، اما این رویکرد نیز همچنان به مرحله استخراج ویژگی دستی وابسته بود و از قابلیت یادگیری سلسله‌مراتبی ویژگی‌ها که در شبکه‌های عمیق وجود دارد، بی‌بهره بود.

۲.۲ .انقلاب یادگیری عمیق

نقطه عطف در حوزه بینایی ماشین، ظهور یادگیری عمیق و شبکه‌های عصبی کانولوشنی بود. این شبکه‌ها، با الهام از ساختار قشر بینایی مغز، قادرند الگوهای فضایی را در تصاویر به‌صورت سلسله‌مراتبی یاد بگیرند. بااین‌حال، آموزش شبکه‌های بسیار عمیق با چالش‌هایی مانند مشکل «محوشدگی گرادیان» [۱۷] مواجه بود که مانع از همگرایی صحیح مدل می‌شد. در این میان، دو معماری نقش کلیدی در حل این مشکلات و پیشبرد حوزه ایفا کردند:

[14] Discrete Cosine Transform (DCT)
[15] Multi-Layer Perceptron (MLP)
[16] Moment Invariants
[17] Vanishing Gradient Problem

معماری ResNet: این معماری که توسط He و همکاران [6] معرفی شد، یک راه‌حل هوشمندانه برای مشکل محوشدگی گرادیان ارائه داد ResNet .با معرفی «اتصالات میان‌بر»[18] یا «بلوک‌های باقی‌مانده»[19] به گرادیان‌ها اجازه می‌دهد تا به راحتی در طول شبکه جریان یابند و بدین ترتیب، آموزش شبکه‌هایی با عمق صدها و حتی هزاران لایه ممکن گردید. ایده اصلی این است که به جای یادگیری یک نگاشت مستقیم (H(x)) ، لایه یاد می‌گیرد که تابع باقی‌مانده (F(x) = H(x) - x) را تخمین بزند. این نوآوری، ResNet را به یکی از پراستفاده‌ترین و موفق‌ترین معماری‌ها در وظایف مختلف بینایی ماشین تبدیل کرد.

معماری EfficientNet: این معماری که توسط Tan و Le [۸] ارائه شد، رویکردی نوین برای مقیاس‌بندی شبکه‌های عصبی معرفی کرد. به جای افزایش بی‌رویه یک بعد از شبکه (عمق، عرض یا رزولوشن)، EfficientNet از یک روش «مقیاس‌بندی ترکیبی»[20] استفاده می‌کند که به طور متوازن هر سه بعد را با استفاده از یک ضریب ثابت افزایش می‌دهد. این کار باعث می‌شود که شبکه با تعداد پارامترها و محاسبات بسیار کمتر، به دقت بالاتری دست یابد و از این رو، "کارآمد" نامیده می‌شود.

۲.۳ .یادگیری انتقال و کارهای اخیر

علاوه بر معماری‌های قدرتمند، تکنیک «یادگیری انتقال» نقش بسزایی در موفقیت‌های اخیر یادگیری عمیق داشته است. ایده اصلی این است که یک مدل که بر روی یک مجموعه‌داده بسیار بزرگ و عمومی مانند ImageNet [9] آموزش دیده است، دانش پایه‌ای ارزشمندی در مورد ویژگی‌های بصری عمومی (مانند لبه‌ها، بافت‌ها و اشکال ساده) کسب کرده است. می‌توان از این دانش از پیش آموخته‌شده به عنوان یک نقطه شروع قوی برای یک وظیفه جدید و تخصصی‌تر (مانند تشخیص حروف فارسی) استفاده کرد. این کار نه تنها زمان آموزش را کاهش می‌دهد، بلکه معمولاً منجر به دستیابی به دقت بالاتر و تعمیم‌پذیری بهتر مدل نیز می‌شود [10] .

باتوجه‌به مرور ادبیات تحقیق، مشخص می‌شود که اگرچه تلاش‌های ارزشمندی در زمینه OCR فارسی صورت‌گرفته است، اما استفاده از معماری‌های ترکیبی که نقاط قوت مدل‌های مختلف را با هم ادغام کنند، کمتر موردبررسی قرار گرفته است. پژوهش حاضر باهدف پر کردن این شکاف، یک راهکار کامل مبتنی بر ترکیب دو معماری پیشرفته را ارائه می‌دهد.

۳ .روش تحقیق

این بخش به روش مورداستفاده در این پژوهش را شرح می‌دهد که شامل سه‌گام اصلی است: (۱) ایجاد و آماده‌سازی مجموعه داده، (۲) پیش‌پردازش تصاویر، و (۳) طراحی و آموزش مدل یادگیری عمیق.

[18] Shortcut/Skip Connections
[19] Residual Blocks
[20] Compound Scaling

۳.۱. مجموعه داده

یکی از ارکان اصلی موفقیت هر مدل یادگیری عمیق، دسترسی به یک مجموعه‌داده بزرگ، متنوع و باکیفیت است. برای این پژوهش، به‌جای اتکا به یک منبع واحد، یک مجموعه‌داده ترکیبی و جامع از دو دیتاست معتبر ایرانی در زمینه حروف دست‌نویس فارسی ایجاد گردید. منابع اصلی مورداستفاده عبارت‌اند از:

۱. مجموعه‌داده معرفی‌شده توسط صدری و همکاران: این مجموعه‌داده یکی از منابع جامع و پراستفاده برای حروف و ارقام دست‌نویس فارسی است[11] .

۲. مجموعه‌داده خیام: این دیتاست جدیدتر که توسط جعفرزاده و همکاران ارائه شده، باهدف پوشش‌دادن تنوع بیشتری از دستخط‌ها و سبک‌های نگارشی ایجاد گردیده است [12] .

با ترکیب و تجمیع نمونه‌ها از این دو منبع، یک مجموعه‌داده نهایی با تقریباً 124 هزار تصویر در 120 کلاس مختلف (شامل اشکال مختلف حروف الفبا در حالات آغازی، میانی، پایانی و تنها) ساخته شد. این حجم و تنوع بالا، بستر مناسبی را برای آموزش یک مدل قوی و قابل تعمیم فراهم می‌آورد. مجموعه داده به سه بخش استاندارد آموزشی (80%)، اعتبارسنجی (10%) و آزمون (10%) تقسیم شد تا ارزیابی مدل به صورت کاملاً استاندارد انجام شود.

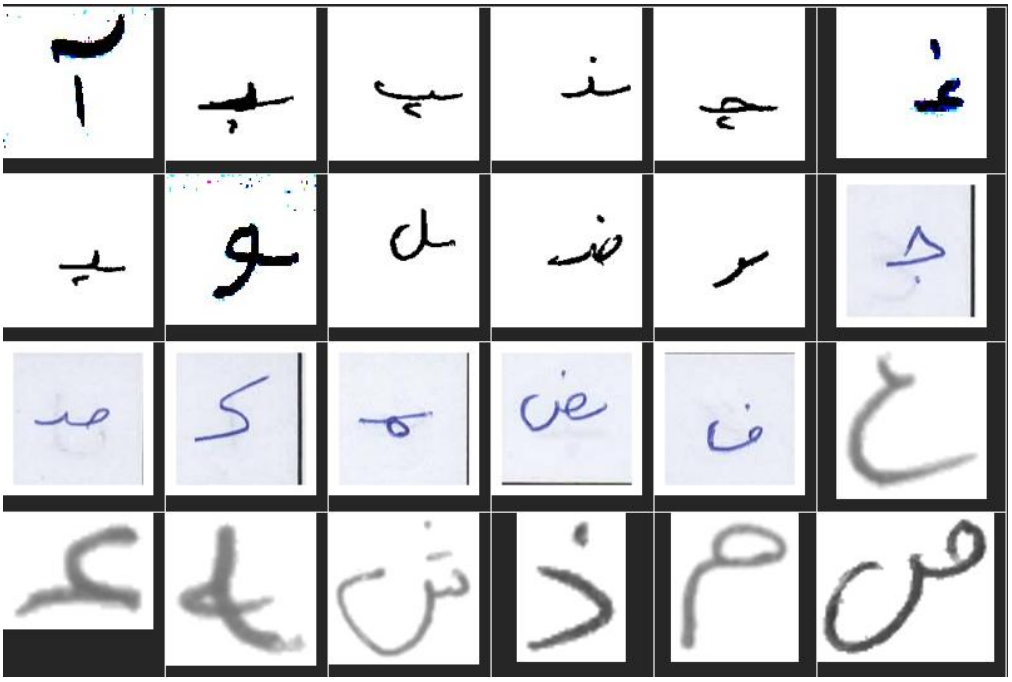
۳. ۲. پیش‌پردازش تصاویر

تصاویر خام موجود در مجموعه‌داده‌ها دارای ابعاد، روشنایی و نویزهای متفاوتی هستند. برای اینکه مدل بتواند الگوهای معنادار را به‌درستی یاد بگیرد، یک مرحله پیش‌پردازش دقیق و استاندارد بر روی تمام تصاویر اعمال گردید. این فرایند که با استفاده از کتابخانه‌های OpenCV [13] و Pillow در پایتون پیاده‌سازی شده، شامل مراحل زیر است :

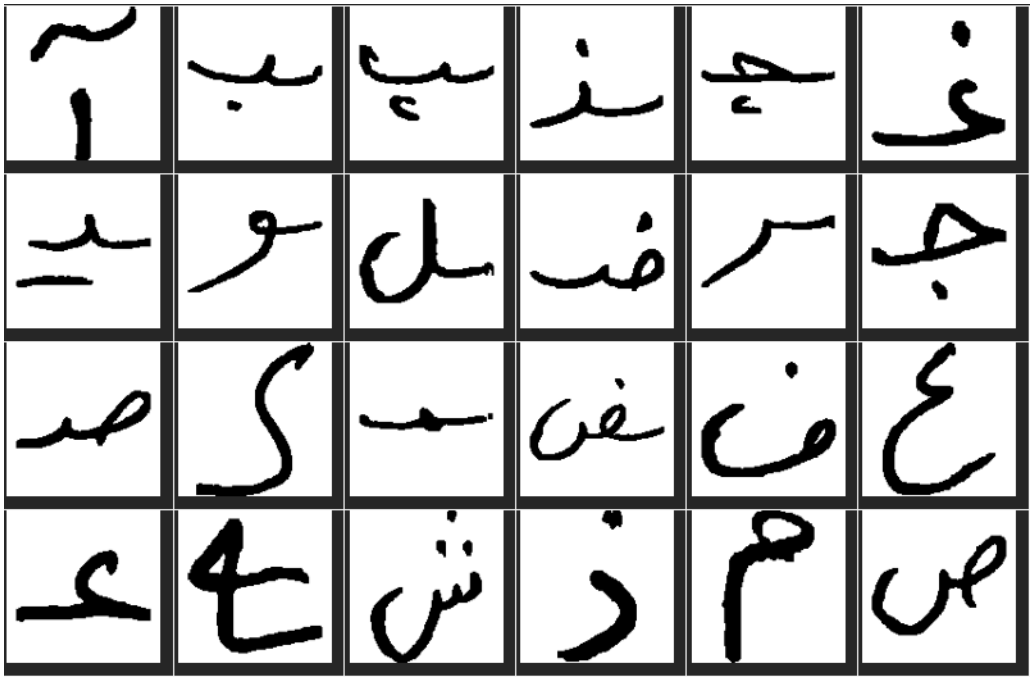
- تبدیل به مقیاس خاکستری :اطلاعات رنگی برای تشخیص حروف ضروری نیست، لذا برای کاهش پیچیدگی محاسباتی، تمام تصاویر به حالت خاکستری تبدیل شدند.
- آستانه‌گذاری دو حالتی (Binarization) :با استفاده از روش آستانه‌گذاری خودکار اتسو [14]، تصاویر خاکستری به تصاویر سیاه‌وسفید تبدیل شدند. این روش به‌صورت هوشمند یک آستانه بهینه برای جداکردن پیش‌زمینه (حرف) از پس‌زمینه پیدا می‌کند.
- برش هوشمند :برای حذف فضاهای خالی و غیرضروری اطراف حرف، ابتدا کانتورهای موجود در تصویر باینری استخراج شده و سپس کوچک‌ترین مستطیل دربرگیرنده[۲۱] حرف پیدا می‌شود. تصویر بر اساس این کادر برش می‌خورد.
- تغییر اندازه با حفظ نسبت ابعاد :تصاویر برش خورده دارای ابعاد متفاوتی هستند. برای استانداردسازی، تمام تصاویر به‌گونه‌ای تغییر اندازه داده می‌شوند که نسبت ابعاد اصلی تصویر حفظ گردد.
- ضخیم‌سازی (Dilation) :برای اطمینان از پیوستگی خطوط و خوانایی بهتر حروف نازک، یک عملیات ضخیم‌سازی با یک کرنل کوچک[۲۲] بر روی تصویر اعمال می‌شود.

[21] Bounding Box
[22] Kernel

- ایجاد حاشیه و مرکزگرایی (Padding) :در نهایت، تصویر تغییراندازه‌یافته در مرکز یک بوم (canvas) مربعی با ابعاد ثابت (مثلاً 128×128 پیکسل) قرار می‌گیرد. این کار تضمین می‌کند که تمام تصاویر ورودی به مدل دارای ابعاد یکسان بوده و حرف در مرکز آن قرار دارد. نتیجه در تصویر ۱ و ۲ مشخص است.
-



شکل ۱ – نمونه ای از تصاویر دیتاست قبل از پیش پردازش



شکل ۲ – نمونه ای از تصاویر دیتاست بعد از پیش پردازش

۳.۳. معماری مدل پیشنهادی

قلب این پژوهش، یک معماری ترکیبی جدید است که از نقاط قوت دو مدل بسیار موفق، یعنی EfficientNet-B0 و ResNet-34، به‌صورت هم‌زمان بهره می‌برد. این معماری به‌گونه‌ای طراحی شده است که دو مسیر موازی برای استخراج ویژگی داشته باشد و سپس اطلاعات این دو مسیر را برای تصمیم‌گیری نهایی ترکیب کند.

- مسیر اول: استخراج‌کننده ویژگی EfficientNet-B0 :اولین ستون فقرات مدل، یک شبکه EfficientNet-B0 است که با وزن‌های از پیش آموزش‌دیده روی ImageNet [9] بارگذاری می‌شود. لایه طبقه‌بند نهایی این شبکه حذف شده و خروجی آن یک بردار ویژگی است که الگوهای استخراج‌شده توسط این معماری کارآمد را نمایش می‌دهد.
- مسیر دوم: استخراج‌کننده ویژگی ResNet-34 :دومین ستون فقرات مدل، یک شبکه ResNet-34 است که آن هم با وزن‌های از پیش آموزش‌دیده ImageNet [9] مقداردهی اولیه شده است. مشابه مسیر اول، لایه طبقه‌بند نهایی این شبکه نیز حذف می‌شود تا خروجی آن یک بردار ویژگی دیگر باشد.
- الحاق ویژگی‌ها (Feature Concatenation) :در این مرحله کلیدی، بردارهای ویژگی که از دو مسیر موازی به دست آمده‌اند، به یکدیگر متصل (الحاق) می‌شوند. این کار یک بردار ویژگی ترکیبی و غنی‌تر ایجاد می‌کند که حاوی اطلاعات استخراج‌شده توسط هر دو معماری است. ایده اصلی این است که هر معماری ممکن است به انواع مختلفی از الگوها حساس باشد و ترکیب آن‌ها به مدل اجازه می‌دهد تا درک جامع‌تری از تصویر ورودی پیدا کند.
- **سر طبقه‌بند** (Classifier Head) **:**بردار ویژگی ترکیبی به یک «سر» طبقه‌بند جدید وارد می‌شود. این سر، یک شبکه عصبی کوچک و تماماً متصل (Fully Connected) است که وظیفه نهایی را بر عهده دارد. این بخش شامل چندلایه خطی، توابع فعال‌ساز ReLU ، لایه‌های نرمال‌سازی دسته‌ای[۲۳] و لایه‌های حذف (Dropout) برای جلوگیری از بیش برازش است. خروجی نهایی این سر، یک بردار با 120 عنصر است که احتمال تعلق تصویر ورودی به هر یک از 120 کلاس حروف را نشان می‌دهد.

۳.۴. فرآیند آموزش

فرایند آموزش مدل با استفاده از کتابخانه PyTorch [15] و با بهره‌گیری از چندین تکنیک پیشرفته برای بهبود عملکرد و پایداری انجام شد.

- افزایش داده (Data Augmentation) :برای جلوگیری از بیش‌برازش و افزایش تنوع داده‌های آموزشی، مجموعه‌ای از تبدیلات تصادفی به‌صورت آنلاین بر روی تصاویر اعمال شد. این تبدیلات شامل چرخش‌های جزئی، جابه‌جایی، تغییر مقیاس، برش پرسپکتیو و پاک‌کردن تصادفی بخشی از تصویر بود[10] .
- هایپرپارامترها و بهینه‌سازی: مدل با استفاده از بهینه‌ساز Adam [16] و تابع هزینه «آنتروپی متقاطع» [۲۴] آموزش داده شد. همچنین از تکنیک «هموارسازی برچسب» [۲۵] برای جلوگیری از اطمینان بیش از حد مدل استفاده گردید.

[23] Batch Normalization
[24] Cross-Entropy Loss

برای مدیریت فرایند آموزش، دو مکانیزم کنترلی به کار گرفته شد :

۱ .زمان‌بند نرخ یادگیری (Learning Rate Scheduler) :از ReduceLROnPlateau استفاده شد که در صورت عدم بهبود هزینه اعتبارسنجی برای چند دوره متوالی، نرخ یادگیری را به‌صورت خودکار کاهش می‌دهد .

۲ .توقف زودهنگام (Early Stopping) :اگر دقت مدل روی داده‌های اعتبارسنجی برای تعداد مشخصی دوره بهبود نمی‌یافت، فرایند آموزش متوقف می‌شد تا از اتلاف منابع و بیش‌برازش جلوگیری شود. هایپرپارامترها و تنظیمات کلیدی در جدول شماره ۱ آمده است.

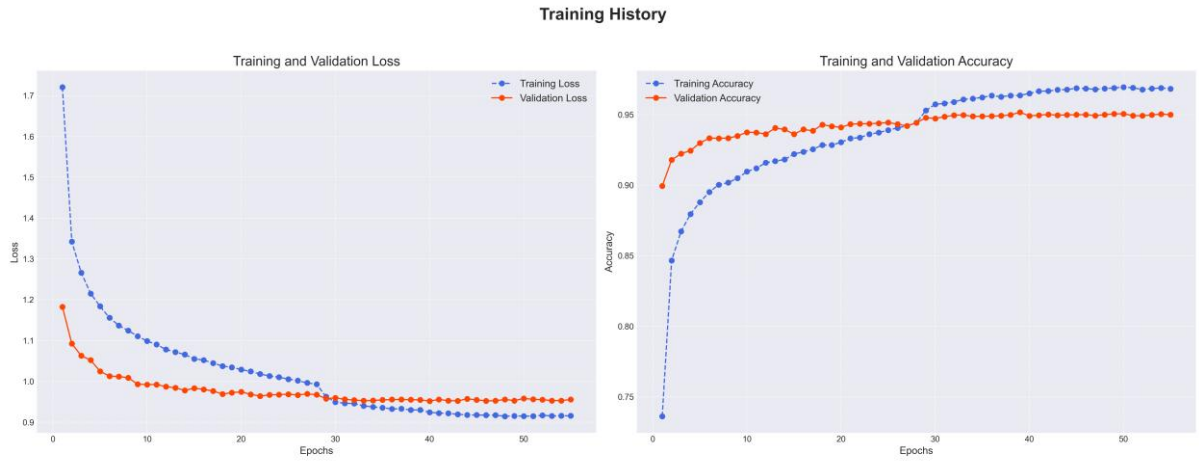
جدول ۱- هایپرپارامترها و تنظیمات کلیدی فرآیند آموزش

پارامتر	مقدار / توضیح
معماری پایه	ترکیبی(EfficientNet-B0 + ResNet-34)
وزن‌های اولیه	از پیش آموزش‌دیده بر رویImageNet
اندازه تصویر ورودی	۳×۲۲۴×۲۲۴
بهینه‌ساز(Optimizer)	Adam
نرخ یادگیری پایه(Backbones)	e-4۱
نرخ یادگیری سر(Head)	1e-3
تابع هزینه (Loss Function)	Cross-Entropyبا Label Smoothing
اندازه دسته(Batch Size)	۳۲
تعداد دوره‌ها(Epochs)	۱۰۰ (با قابلیت توقف زودهنگام)
زمان‌بند نرخ یادگیری	ReduceLROnPlateau (patience = 5)
توقف زودهنگام	Patience = 15

[25] Label Smoothing

۴ .یافته‌ها و تحلیل نتایج

در این بخش، نتایج کمی و کیفی حاصل از آموزش و ارزیابی مدل پیشنهادی ارائه می‌گردد. مدل بر روی یک سیستم مجهز به پردازنده گرافیکی (GPU) آموزش داده شد تا فرایند محاسباتی تسریع گردد. نتیجهٔ این آموزش در نمودار شماره ۱ مشخص است.



نمودار ۱ – روند تغییرات هزینه (چپ) و دقت (راست) در طول دوره‌های آموزش

نمودار ۱، روند تغییرات دقت و هزینه را بر روی مجموعه داده‌های آموزشی و اعتبارسنجی در طول دوره‌های آموزش نشان می‌دهد. تحلیل این نمودار چندین نکته مهم را آشکار می‌سازد:

- تحلیل منحنی‌های هزینه (نمودار چپ) :همان‌طور که مشاهده می‌شود، هزینه آموزش (خط آبی) و هزینه اعتبارسنجی (خط نارنجی) هر دو با شیب تندی در دوره‌های ابتدایی کاهش می‌یابند که نشان‌دهنده یادگیری سریع مدل در مراحل اولیه است. پس از حدود ۳۰ دوره، شیب کاهش ملایم‌تر شده و منحنی‌ها به سمت همگرایی پیش می‌روند. وجود یک شکاف کوچک و پایدار بین منحنی‌های آموزش و اعتبارسنجی نشان می‌دهد که مکانیزم‌های تنظیم‌گری مانند افزایش داده و Dropout به‌خوبی توانسته‌اند از بیش‌برازش شدید جلوگیری کنند. مدل در نهایت به کمینه هزینه اعتبارسنجی ۰.۹۵۰۶ دست می‌یابد.
- تحلیل منحنی‌های دقت (نمودار راست) :منحنی‌های دقت نیز رفتار مشابهی را نشان می‌دهند. دقت آموزش و اعتبارسنجی هر دو به‌سرعت افزایش‌یافته و به یک سطح پایدار نزدیک می‌شوند. یک نکته جالب در نمودار دقت، جهش‌های کوچک در عملکرد است که می‌تواند به دلیل فعال‌شدن زمان‌بند نرخ یادگیری و خروج مدل از بهینه‌های محلی (local minima) باشد. مدل در نهایت به بیشینه دقت ٪۹۵.۰۶ بر روی داده‌های اعتبارسنجی دست می‌یابد که نشان‌دهنده قدرت بالای مدل در تعمیم آموخته‌های خود به داده‌های دیده‌نشده است.

این نتایج کمی نشان می‌دهند که معماری ترکیبی پیشنهادی به همراه تکنیک‌های پیشرفته آموزش، توانسته است یک مدل قدرتمند و قابل‌اعتماد برای بازشناسی حروف فارسی ایجاد کند.

۵ .بحث و نتیجه‌گیری

در این پژوهش، یک سامانه جامع و کارآمد برای بازشناسی حروف دست‌نویس فارسی با استفاده از یک معماری ترکیبی مبتنی بر یادگیری عمیق طراحی و پیاده‌سازی شد. نتایج به‌دست‌آمده که حاکی از دستیابی به‌دقت ٪۹۵.۰۶ بر روی مجموعه‌داده اعتبارسنجی است، چندین نکته مهم را برجسته می‌سازد:

۱ .**کارایی معماری ترکیبی :**عملکرد بالای مدل پیشنهادی، کارایی فوق‌العاده ترکیب دو معماری قدرتمند و متفاوت EfficientNet و ResNet را تأیید می‌کند. این رویکرد به مدل اجازه می‌دهد تا از الگوهای بصری متنوع‌تری که توسط هر شبکه به‌صورت مستقل استخراج می‌شود، بهره‌مند گردد و به درک عمیق‌تری از داده‌ها دست یابد.

۲ .**اهمیت یادگیری انتقال :**موفقیت این پژوهش به شدت به تکنیک یادگیری انتقال وابسته است. با بهره‌گیری از دانش نهفته در وزن‌های از پیش آموزش‌دیده بر روی دیتاست ImageNet [9] ، مدل توانست با داده‌های نسبتاً کمتر و در زمان کوتاه‌تر به همگرایی مطلوبی دست یابد. این موضوع نشان می‌دهد که ویژگی‌های بصری پایه‌ای که در لایه‌های اولیه شبکه‌های عصبی یاد گرفته می‌شوند، تا حد زیادی بین دامنه‌های مختلف تصویری قابل انتقال هستند.

۳ .**نقش کلیدی پیش‌پردازش و افزایش داده :**الگوریتم توسعه داده شده برای نرمال‌سازی تصاویر، با حذف نویز، استانداردسازی ابعاد و مرکزگرایی حروف، نقش کلیدی در آماده‌سازی داده‌ها برای یادگیری مؤثر توسط شبکه عصبی ایفا کرد. همچنین، استفاده از تکنیک‌های افزایش داده برای مقاوم‌سازی مدل در برابر تنوع دست‌خط‌ها و جلوگیری از بیش‌برازش حیاتی بود.

باوجود نتایج امیدوارکننده، این پژوهش دارای محدودیت‌هایی نیز است. مدل حاضر بر روی تشخیص حروف به‌صورت مجزا آموزش‌دیده است و قابلیت تشخیص کلمات یا جملات پیوسته را ندارد.

۶ .پیشنهادات برای کارهای آتی

بر اساس یافته‌ها و محدودیت‌های این پژوهش، موارد زیر برای تحقیقات آینده پیشنهاد می‌گردد:

- توسعه برای تشخیص کلمات :می‌توان از معماری‌های پیشرفته‌تری مانند شبکه‌های عصبی بازگشتی کانولوشنی[۲۶] (CRNN)یا مدل‌های مبتنی بر ترنسفورمر[۲۷] (Transformer) استفاده کرد که قادر به پردازش توالی‌ها هستند و می‌توانند کلمات کامل را بدون نیاز به قطعه‌بندی حروف، بازشناسی کنند.
- گسترش مجموعه‌داده :باوجود جامع بودن دیتاست مورداستفاده، افزودن نمونه‌های بیشتر از دست‌خط‌های متنوع‌تر، به‌خصوص دست‌خط‌های محاوره‌ای و شکسته، می‌تواند به افزایش مقاومت و دقت مدل در شرایط واقعی کمک کند.
- بهینه‌سازی مدل برای اجرا بر روی دستگاه‌های با منابع محدود :مدل ترکیبی حاضر نسبتاً بزرگ است. می‌توان با استفاده از تکنیک‌های فشرده‌سازی مدل مانند هرس‌کردن[۲۸] (Pruning) یا کوانتیزه‌سازی[۲۹](Quantization) ، یک نسخه سبک‌تر از مدل را برای استفاده در تلفن‌های همراه یا سایر دستگاه‌های لبه[۳۰] ایجاد کرد.

[26] Convolutional Recurrent Neural Network
[27] Transformer Models
[28] Pruning
[29] Quantization

منابع

۱. اسدی، فرشید؛ صیدی پیری، رسول؛ نوری، زینب؛ لطفی پور، امین. (۱۳۹۵). تشخیص حروف دست‌نویس فارسی با استفاده از حسگر شتاب‌سنج و الگوریتم‌های یادگیری ماشین .چهارمین کنفرانس بین‌المللی در مهندسی برق و کامپیوتر.

۲. مدحتی، امید؛ کریمی، حسین؛ قائدی، فرشاد؛ طاهریان، محسن. (۱۳۹۴). تشخیص حروف دست‌نوشته فارسی با استفاده از درخت تصمیم باینری .سومین همایش ملی کامپیوتر.

۳. توپچی، مهدی؛ ابوالقاسم‌پور، سیده اعظم. (۲۰۱۷). استفاده از شبکه عصبی پرسپترون چندلایه جهت تشخیص حروف الفبای فارسی .دومین کنفرانس بین‌المللی یافته‌های نوین پژوهشی در علوم، مهندسی و فناوری

4. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
5. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
7. Bonyani, M., Jahangard, S., & Daneshmand, M. (2020). Persian Handwritten Digit, Character, and Words Recognition by Using Deep Learning Methods. arXiv preprint arXiv:2010.12880.
8. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105-6114). PMLR.
9. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).
10. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1-48.
11. Sadri, J., Yeganehzad, M. R., & Saghi, J. (2016). A novel comprehensive database for offline Persian handwriting recognition. Pattern Recognition, 60, 378-393.
12. Jafarzadeh, Pourya, Padideh Choobdar, and Vahid Mohammadi Safarzadeh. "Khayyam Offline Persian Handwriting Dataset." arXiv preprint arXiv:2406.01025 (2024)
13. Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
14. Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.
15. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G.,... & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.
16. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[30] Edge Devices

Recognition of Persian Handwritten Characters Using a Deep Hybrid Model Based on ResNet and EfficientNet Architectures

Mohammad Matin Mohammadi Pega[1], Omid Taheri[2]

[1] Student of Computer, National University of Skill (NUS), Tehran, Iran, matinmmp1381@gmail.com.
[2] Department of Computer, National University of Skill (NUS), Tehran, Iran, omid.taheri6080@gmail.com.

Abstract— Optical Character Recognition (OCR) for the Persian language has always been a challenging area in artificial intelligence due to the structural complexities of the script. In recent years, deep learning models, especially Convolutional Neural Networks (CNNs), have made significant progress in this field. This research presents a comprehensive system for recognizing handwritten Persian characters using a modern and hybrid deep learning architecture. For this purpose, a large dataset containing 124,000 images from 120 different classes of Persian characters was created by combining two standard datasets. The images were then normalized using a custom pre-processing algorithm. In this study, we have built a hybrid architecture based on two well-known networks, **EfficientNet** and **ResNet-34**, using the transfer learning technique. To enhance the model's resistance to overfitting, techniques such as data augmentation and early stopping were employed. The evaluation results show that the proposed model achieved a maximum accuracy of **95.06%** on the validation data. This high performance demonstrates the effectiveness of modern deep learning approaches in solving the complex problem of Persian character recognition.

Keywords: Optical Character Recognition, Deep Learning, Convolutional Neural Networks, Transfer Learning, Hybrid Architecture, Persian Characters.