مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال دهم، شماره دوم، تابستان ۱۴۰۲

معرفی یک مجموعه اسناد متنی فارسی برای کاربردهای درک و ناحیه‌بندی اسناد فارسی

امین فرجی[۱]، مسعود سعید[۲] و حسین نظام آبادی‌پور[۳]

چکیده

وجود مجموعه‌داده‌های تصویری نقش اساسی در زمینه تشخیص نویسه‌خوان نوری (OCR) و بازیابی اسناد دارد. علی‌رغم اینکه تا به امروز مجموعه‌داده‌های تصویری زیادی با اشیا متفاوت در حوزه درک و ناحیه‌بندی اسناد غیرفارسی منتشر شده است،رسم‌الخط فارسی از این پیشرفت عقب‌مانده است و تاکنون در زمینه درک و ناحیه‌بندی اسناد فارسی، مجموعه‌دادگانی با دسترسی عمومی ارائه نشده است. از سوی دیگر، اگرچه زبان‌های فارسی و عربی شباهت‌های زیادی دارند، اما تفاوت بین ساختار این دو زبان باعث می‌شود که سیستم‌های آموزش‌دیده OCR با مجموعه‌داده عربی، دقت مناسبی روی تصاویر اسناد فارسی نداشته باشند. در این مقاله،یک مجموعه‌داده برای تصاویر اسناد فارسی معرفی می‌گردد که مشتمل بر ۵۵۹۸ تصویر است. تصاویر تهیه شده متعلق به روزنامه‌ها، کتاب‌های درسی،مقالات علمی، فایل‌های PDF فارسی، پایان‌نامه‌ها، انواع لوگو ایرانی، کتب دست‌نوشته قدیمی و جزوات تایپ شده و دست‌نویس ریاضی هستند. در مجموعه داده معرفی شده، اشیا درون تصاویر به ۶ گروه پاراگراف (متن)، شکل، جدول، لوگو، رابطه ریاضی و سرصفحه‌دسته‌بندیو برچسب‌گذاری شده‌اند. برای ارزیابی کارایی مجموعه تصویر پیشنهادی، سه روش شناخته شده مبتنی بر یادگیری عمیق پیاده سازی و نتایج بر مبنای معیارهای مختلف گزارش شده است.

کلید واژه ها

پردازش تصویر،اسناد متن فارسی، ناحیه‌بندیسند، درک سند،مجموعه داده.

۱ مقدمه

ناحیه‌بندی تصویر به فرآیند افراز یک تصویر دیجیتال به نواحی سازنده آن، گفته می‌شود. هدف از ناحیه‌بندی تصویر، ساده‌سازی یا تغییر در نمایش یک تصویر به‌گونه‌ای است که هم معنی‌دارتر و هم برای تحلیل آسان‌تر باشد. به طور معمول، از ناحیه‌بندی تصویر برای پیداکردن نواحی مورد نظر (خطوط، منحنی‌ها و غیره) در تصویر استفاده می‌شود. از کاربردهای مهم ناحیه‌بندی تصویر، می‌توان به درک و ناحیه‌بندی تصویر اسناد تاریخی[۱]، مجلات [۲, ۳]و مقالات علمی [۴, ۵] اشاره کرد که در فرمت تصویر دیجیتال ارائه می‌شوند.

افزایش سریع دیجیتالی‌شدن تصاویر اسناد به طور قابل توجهی دسترسی به داده‌ها را بهبود بخشیده است. استخراج دستی اطلاعات از تصاویر اسناد پویش شده یک فرآیند دشوار، زمان‌بر و گاهی غیرممکن است؛ بنابراین استخراج اطلاعات از تصویر اسناد به یک زمینه تحقیقاتی مهم در بین پژوهشگران تبدیل شده است. برخی از اطلاعات درون اسناد به صورت گرافیکی مانند شکل، جدول و رابطه‌ها ذخیره می‌شوند که در حوزه درک و ناحیه‌بندی اسناد به عنوان اطلاعات گرافیکی درون اسناد [۶] به آن‌ها اشاره می‌شود. جدا از اطلاعات گرافیکی، استخراج متن از تصویر اسناد با استفاده از روش‌های تشخیص نویسه‌خوان نوری هنوز یک موضوع مهم برای تحقیق است. از جمله کاربردهای تشخیص

این مقاله در تیرماه سال ۱۴۰۱ دریافت شد؛ در آبان‌ماه بازنگری و در آذرماه همان سال پذیرفته شد.

[۱] دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی، دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان، کرمان، ایران
رایانامه: aminfaraji1000@eng.uk.ac.ir
[۲] دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان، کرمان، ایران
رایانامه: msaeedmz@uk.ac.ir
[۳] دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان، کرمان، ایران
رایانامه: nezam@uk.ac.ir

نویسنده مسئول: مسعود سعید

dorl.net/dor/20.1001.1.23831197.1402.10.2.3.3

نویسه خوان نوری می‌توان به بازشناسی دست نوشته‌ها [۷]، تشخیص خودکار پلاک خودرو [۸]، خواندن چک‌های بانکی و اعتبار سنجی امضاء افراد [۹] اشاره کرد. با اینکه روش‌های مختلفی برای استخراج ویژگی‌های متنی درون تصویر اسناد پیشنهاد شده است، اما برخی از این روش‌ها در استخراج ویژگی متنی درون گروه های شکل، جدول و رابطه‌ها دقیق و کارآمد نیستند[۱۰].

یکی از مهم‌ترین روش‌های استخراج اطلاعات درون تصویر، استفاده از شبکه‌های عصبی است. این روش‌ها در مقایسه با روش‌های سنتی یادگیری ماشین به نتایج قابل قبولی دست پیدا کرده‌اند. از مهم‌ترین الگوریتم‌های شبکه‌های عصبی، می‌توان به شبکه‌های عصبی کانولوشنی اشاره کرد. شبکه عصبی کانولوشن نوع خاصی از شبکه عصبی عمیق است که داده‌هایی را که آرایش شبکه‌ای دارند، پردازش کرده و سپس ویژگی‌های مهم آن‌ها را استخراج می‌کند. یکی از مهم‌ترین چالش‌های شبکه عصبی کانولوشنی کمبود دادگان کافی برای آموزش آن است. در این تحقیق ما به دنبال ناحیه‌بندی تصاویر[۱] اسناد فارسی با کمک شبکه‌های عصبی کانولوشنی هستیم که متأسفانه برای انجام این تحقیق با مشکل نبود یک مجموعه داده استاندارد با حجم کافی از تصاویر اسناد فارسی روبرو شدیم. از این رو، بر آن شدیم تا در ابتدا یک مجموعه داده از تصویر اسناد فارسی فراهم نماییم. در این راستا، در گام اول یک مجموعه داده برای درک و ناحیه‌بندی روی اسناد فارسی معرفی شده و در ادامه برای ارزیابی این مجموعه داده از سه شبکه کانولوشنی شناخته شده که در حوزه ناحیه‌بندی تصویر دیجیتال ارائه شده است، استفاده می‌شود.

سایر بخش‌های این مقاله به این ترتیب سازمان‌دهی شده‌اند که در بخش دوم به پیشینه پژوهشی پرداخته می‌شود. در این بخش به معرفی مجموعه داده‌های ارائه شده روی اسناد انگلیسی، چینی، عربی و دیگر زبان‌ها می‌پردازیم. در بخش سوم مجموعه داده پیشنهادی به همراه ویژگی‌های آنها ارائه می‌شود. در بخش چهارم انواع معیارهای ارزیابی شباهت در حوزه درک و ناحیه‌بندی تصویر معرفی می‌شوند. در بخش پنجم به معرفی مدل‌های کانولوشنی پرداخته می‌شود. در بخش ششم نتایج و خروجی استفاده از مدل کانولوشنی آورده شده و نهایتاً در بخش هفتم به جمع‌بندی مقاله پرداخته می‌شود.

۲ پیشینه پژوهشی

ناحیه‌بندی تصویر از مباحث مهم در حوزه پردازش تصویر است. رویکردهای ناحیه‌بندی تصویر را می‌توان به سه گروه طبقه‌بندی کرد: (۱) ناحیه‌بندی معنایی[۲]، (۲) ناحیه‌بندی نمونه[۳] و (۳) ناحیه‌بندی پانوپتیک[۴]. رویکرد ناحیه‌بندی معنایی، به فرآیند پیوند دادن هر پیکسل تصویر به یک برچسب گروه مشخص اشاره می‌کند.به عبارت دیگر، با کلیه اشیا موجود در یک تصویر که متعلق به یک گروه هستند به عنوان یک شیء واحد با برچسب همان گروه برخورد می‌شود. در رویکردناحیه‌بندی نمونه، مشابه ناحیه‌بندی معنایی به هر کدام از پیکسل‌های تصویر یک برچسب زده می‌شود، اما با این تفاوت که با هر یک از اشیا یک گروه مشخص به عنوان نمونه‌های جدا از هم برخورد می‌شود. ناحیه‌بندی پانوپتیک از ترکیب رویکرد ناحیه‌بندی معنایی و ناحیه‌بندی نمونه شکل گرفته است و هر پیکسل از تصویر را با دو مقدار نمایش می‌دهد، یکی برچسب گروه و دیگری معرف درصد تعلق پیکسل به هر گروه است.

بیشتر فعالیت‌های صورت گرفته در حوزه درک و ناحیه‌بندی اسناد با استفاده از رویکرد ناحیه‌بندی معنایی ارائه شده است. از مهم‌ترین کاربردهای این رویکرد می‌توان به ناحیه‌بندی تصویر در خودروهای خودران، تصاویر پزشکی، نجوم و اسناد تاریخی اشاره کرد. در این حوزه‌ها با کمک الگوریتم‌هایادگیری‌عمیقشاهد پیشرفت چشمگیری هستیم [۱۵, ۱۴, ۱۳, ۱۲, ۱۱] که این پیشرفت‌ها به واسطه در دسترس بودن مجموعه داده‌های عمومی امکان‌پذیر شده است [۱۹, ۱۸, ۱۷, ۸, ۱۶]. از مهم‌ترین چالش‌های به کارگیری الگوریتم‌های یادگیری عمیق در این حوزه می توان به مجموعه داده‌های کوچک و عدم توازن داده‌ها در گروه‌های مختلف اشاره کرد. اغلب مجموعه داده‌های موجود در حوزه درک و ناحیه‌بندی اسناد،از مشکل تعداد محدود تصویر[۲۱, ۲۰, ۵, ۴] یا تعداد محدود اشیا برچسب خورده در تصویر رنج می‌برند [۴]. عموم این داده‌ها برای یادگیری شبکه‌های عصبی کانولوشنی کافی نیستند. برخی از پژوهش‌های صورت گرفته در این حوزه سعی کرده‌اند که چالش کمبود داده را از طریق تولید تصاویر مصنوعی جبران کنند[۵]. از مجموعه داده‌هایی که برایدرکوناحیه‌بندیاسناد مورد ارائه شده است می توان به مجموعه داده RDCL[۵] اشاره کرد[۲]. این مجموعه داده از مجلات انگلیسی زبان جمع‌آوری شده و متشکل از ۷ تصویربرایآموزش۷۰ تصویر برای آزمون است. در این مجموعه داده،از پاراگراف، عنوان صفحه،شماره صفحه و پاورقی برای گروه‌های هدف استفاده شده است. مجموعه داده دیگر، DSSE_200 است که مشتمل بر ۲۰۰ نمونه تصویر از اسناد، مجلات و اسلایدها به زبان انگلیسی در ۶ گروه پاراگراف،لیست، جدول، شکل،پاورقی و سرصفحه است [۵].

در سال ۲۰۱۴ مجموعه داده Tao [۲۰] به عنوان بزرگترین مجموعه داده در حوزه درک و ناحیه‌بندی اسناد معرفی شد. این مجموعه داده در مجموع شامل ۲۲۴ تصویر و از ۳۵ کتاب الکترونیکی چینی و انگلیسی تهیه شده است. مجموعه داده‌های

[۱] Image Segmentation
[۲] Semantic Segmentation
[۳] Instance Segmentation
[۴] Panoptic Segmentation
[۵] Recognition of Documents with Complex Layouts

CS_Large و CS_150[۴]، مجموعه داده‌هایی به زبان انگلیسی هستند که به ترتیب شامل ۳۱۰۰ و ۱۵۰ تصویر می‌باشند. مجموعه داده CS_150، برای جمع‌آوری تصویر اسناد خود از متن‌های درون مقالات منتشر شده در چندین کنفرانس استفاده کرده است و مجموعه داده CS_large، به صورت تصادفی از مقالاتی که در Scholar Semantic[۱] ارائه شده‌اند برای جمع آوری تصاویر استفاده نموده است. گروه‌های مبتنی بر تصویرهای درون مجموعه داده CS_large عبارت‌اند از: محدوده پاراگراف، شکل، جدول و زیرنویس شکل‌ها. مجموعه داده SPaSe از اسلایدهای آموزشی مدارس و دانشگاه‌ها برای تهیه نمونه‌های مجموعه داده استفاده کرده است [۲۲]. این اسلایدها به زبان فرانسوی، انگلیسی، رومانیایی و ویتنامی تهیه شده است. تعداد نمونه‌های این مجموعه داده شامل ۲۰۰۰ تصویر است که در ۲۵ گروه دسته‌بندی شده است. برخی از گروه‌های این مجموعه داده دارای همپوشانی هستند.

تاکنون مجموعه داده‌های معرفی شده در این حوزه بر روی اسناد غیرفارسی بوده است. علی‌رغم پیشرفت‌های گسترده در ناحیه‌بندی اسناد زبان انگلیسی، سایر زبان‌ها مانند عربی، اردو و به‌ویژه فارسی در این حوزه عقب‌مانده‌اند. مجموعه داده ارائه شده در مرجع[۲۳] متشکل از ۵۰ تصویر از اسناد عربی است. نویسندگان این مقاله، نمونه‌های این مجموعه داده را از وب سایت خبری النهار[۲]، الحیاه[۳] و مجله خبری لبنان[۴] تهیه کرده اند. گروه‌های معرفی شده در این مجموعه داده عبارت‌اند از: قاب (قاب تصویر)، عکس، خط متن و جدول.مجموعه داده مرجع [۲۴] شامل ۳۸ نمونه از تصویر اسناد عربی می‌باشد. این مجموعه داده از ۷ کتاب عربی در دو گروه پاراگراف و متن حاشیه صفحه تشکیل شده است. مجموعه داده تعریف شده در مرجع [۲۵]بزرگترین مجموعه داده اسناد عربی است. این مجموعه داده شامل ۹۰۰۰ تصویر از ۷۰۰ کتاب عربی می‌باشد.

بیشتر فعالیت‌هایی صورت گرفته در حوزه درک و ناحیه‌بندی اسناد بر روی دو گروه خاص مانند (متن و حروف) تمرکز کرده‌اند. مجموعه داده مرجع[۲۶] از ۱۶۳ تصویر برای ناحیه‌بندی حروف عربی استفاده می‌کند. این مجموعه داده مشتمل بر ۱۶۸۰۰ حروف عربی است که توسط ۶۰ نویسنده تهیه شده است. هر نویسنده از حروف (الف) تا (ی) را ۱۰ مرتبه به دو صورت نوشته است که یکی برای تصویر مجموعه داده آموزش و دیگری برای مجموعه داده آزمون مورد استفاده قرار گرفته است. ۴۳ تصویر این مجموعه داده که شامل ۱۳۴۴۰ حروف است برای آموزش و ۱۲۰ تصویر دیگر که شامل ۳۳۶۰ حروف است برای آزمون استفاده می‌شود.

در مجموعه داده مرجع [۲۷] از دست نوشته‌های پتر بزرگ[۵] امپراطور روسیه استفاده شده است و مشتمل بر ۹۶۹۴ تصویر با فایل متنی پویش شده از فرمان‌ها، نامه‌ها و یاداشت‌ها است. از آنجایی که این مجموعه داده برای درک و ناحیه‌بندی خطوط متن پیشنهاد شده، شامل ۲۶۵۷۸۸ حروف و ۵۰۹۹۸ کلمه می‌باشد. در این مجموعه داده، از ۶۲۳۷ تصویر برای آموزش، ۱۹۳۰ تصویر برای اعتبارسنجی و ۱۵۲۷ تصویر دیگر برای مجموعه داده آزمون استفاده شده است.

در مرجع [۲۸] مجموعه داده دیگری متشکل از ۱۰۰۰ تصویر پویش شده به زبان عبری ارائه شده است. این مجموعه داده اولین مجموعه داده به زبان عبری است که برای ناحیه‌بندی حروف، طبقه‌بندی و تشخیص کلمات پیشنهاد شده است.مرجع[۲۹]، با استفاده از روش‌های شناسایی نویسه‌خوان نوری[۶]، به استخراج متن‌های فارسی از سه روزنامه فارسی، دو کتاب الکترونیکی فارسی و روزنامه‌های ورزشی پرداخته است. گروه‌های معرفی شده در این مجموعه داده شامل پاراگراف و متن حاشیه تصاویر است.

مرجع [۳۰] به معرفی یک مجموعه داده با ۱۰ گروه لوگو متفاوت پرداخته است. این مجموعه داده شامل ۲۷۴ تصویر از گروه لوگو است که هر کدام از این لوگوها با ابعاد و رنگ‌های متفاوت نمونه‌برداری شده است. لوگوهای نمونه‌برداری درون این مرجع عبارت‌اند از: لوگو بانک ملی، سپه، تجارت، شرکت‌های ارتباطات، آب، گاز، دانشگاه آزاد اسلامی، پیام‌نور، شریف و علمی کاربردی.

با توجه به اینکه تاکنون مجموعه داده جامعی در حوزه درک و ناحیه‌بندی اسناد فارسی معرفی نشده است، در این تحقیق یک مجموعه داده متشکل از ۶ گروه معرفی شده است. مقایسه این مجموعه داده پیشنهادی با سایر مجموعه داده های شناخته شده در این حوزه در جدول(۱) ارائه شده است. همانطور که مشاهده می‌شود مجموعه داده معرفی شده در مقایسه با برخی از مجموعه داده‌های غیرفارسی از تعداد نمونه‌های بیشتر و از گروه‌های متنوع‌تری برخوردار است.

۳ مجموعه‌داده ناحیه‌بندی اسناد فارسی

اغلب مجموعه داده‌های موجود در حوزه درک و ناحیه‌بندی اسناد، غیرفارسی هستند. در این تحقیق اولین مجموعه داده برای کاربردهای ناحیه‌بندیاسناد فارسی معرفی شده است که متشکل از ۶ گروه می باشد.

[۱] https://github.com/allenai/pdffigures2
[۲] ANNAHAR
[۳] ALHAYAT
[۴] Labanon
[۵] Peter The Great
[۶] Optical Character Recognition (OCR)

هدف از ارائه این مجموعه داده، درک و ناحیه بندی نواحی مورد نظر برای تصاویر اسناد فارسی به صورت تمام خودکار است. در ادامه این بخش، ابتدا توضیح کاملی از فرآیندجمع‌آوری تصاویر و برچسب‌زنی آن‌ها ارائه می‌شود. سپس، تجزیه و تحلیل دقیقی از گروه‌های معرفی شده آورده می‌شود. نهایتاً، در بخش پایانی توضیح جامعی از نویزهای ایجاد شده درون تصویر و چرخش آن‌ها ارائه می‌شود.

جدول (۱):مقایسه ویژگی‌های تعدادی از مجموعه‌های داده موجود و مجموعه داده پیشنهادی در حوزه درک و ناحیه‌بندی اسناد

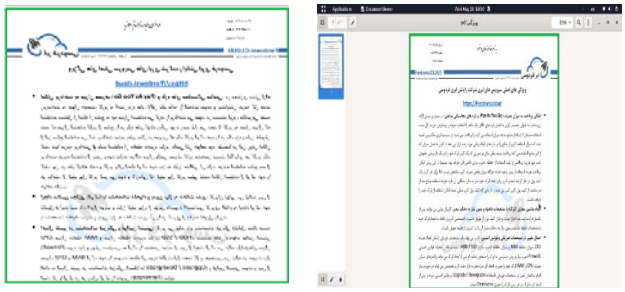
مجموعه داده	تعداد تصویر	جدول	عکس	پاراگراف	لوگو	رابطه ریاضی	عنوان	زبان	سال
RDCL [۲]	۷۷	×	×	✓	×	×	✓	انگلیسی	۲۰۱۵
DSSE-200 [۵]	۲۰۰	✓	✓	✓	×	×	✓	انگلیسی	۲۰۱۷
SpaSe [۲۲]	۲۰۰۰	✓	✓	×	✓	×	✓	انگلیسی	۲۰۱۹
[۲۳]	۵۰	✓	✓	✓	×	×	✓	عربی	۲۰۰۳
[۲۴]	۳۸	×	×	✓	×	×	✓	عربی	۲۰۱۲
BE-Arabic [۲۵]	۹۰۰۰	×	✓	✓	×	✓	✓	عربی	۲۰۲۱
ADIR [۲۶]	۱۶۳	×	×	✓	×	×	✓	عربی	۲۰۲۱
Digital Peter [۲۷]	۹۶۹۴	×	×	✓	×	×	✓	روسیه	۲۰۲۱
HHID [۳۱]	۲۷۰	×	×	✓	×	×	✓	انگلیسی	۲۰۲۲
HHD [۲۸]	۱۰۰۰	×	×	✓	×	×	✓	عبری	۲۰۲۰
PerLogo [۳۰]	۲۷۴	×	×	×	✓	×	×	فارسی	۲۰۱۷
پیشنهادی[۱]	۵۵۹۸	✓	✓	✓	✓	✓	✓	فارسی	۲۰۲۲

۳-۱ جمع‌آوری تصاویر و برچسب‌زنی

مجموعه داده پیشنهاد شده در این تحقیق مشتمل بر ۵۵۹۸تصویر است. نواحی مورد نظر درون تصاویر این مجموعه داده در ۶ گروهدسته‌بندی می‌شوند که عبارت‌اند از: پاراگراف، سرصفحه، لوگو، شکل، رابطه ریاضی و جدول. با هدف ایجاد تنوع در انتخاب تصاویر،برای تهیه این مجموعه داده از کتب درسی در مقاطع دبستان و دبیرستان، فایل‌های PDF فارسی، مقالات مرکز اطلاعات علمی و جهاد دانشگاهی[۱]،کتاب‌های دست‌نویس قدیمی، لوگوهای ایرانی، لوگوهای شبکه‌های اجتماعی، پایان‌نامه‌ها و روزنامه‌های فارسی زبان استفاده شده است. تصاویر این مجموعه داده دارای اندازه های متفاوتی هستند. بزرگترین اندازه تصویر ۲۳۳۹×۱۶۵۴ پیکسل در پیکسل و کمترین آن ۳۱۹×۳۱۹ پیکسل در پیکسل می‌باشد. منابع استفاده شده جهت نمونه برداری و تعداد تصاویر به شرح زیر می باشند:

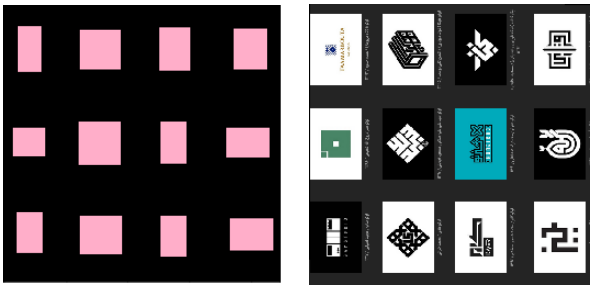
- ۱۰۹۷ تصویر از روزنامه‌ها و مجلات نمونه برداری شده است که حاوی اطلاعات پاراگراف، سرصفحه، لوگو، جدول و شکل هستند.
- ۲۳ تصویر مشتمل بر ۳۴۵ رابطه ریاضی تایپ شده نمونه برداری شده است.
- ۵۱ تصویر مشتمل بر ۴۷۳ رابطه ریاضی تایپ شده که به صورت دستی نمونه برداری شده است.
- ۴۰۷ تصویر که از ۶ پایان‌نامه در رشته‌های مختلف نمونه برداری شده است.
- ۴۳ تصویر از کتاب دست‌نویس قدیمی منتظم ناصری نمونه برداری شده است.
- ۵۰ تصویر از کتاب دست‌نویس عجائب المخلوقات (شگفتی‌های آفریدگان و موجودات)نوشته زکریا قزوینی جمع آوری شده است.
- ۳۳ تصویر لوگو که از شبکه‌های اجتماعی غیرفارسی جمع‌آوری شده است.
- ۳۱۲تصویر لوگو ایرانی با ابعاد مختلف که از شبکه‌های اجتماعی فارسی جمع‌آوری شده است.
- ۱۲ تصویر از یک مقاله در حوزه پزشکی با عنوان "بررسی تأثیر نانوتکنولوژی بر علوم پزشکی و زیست محیطی از دیدگاه ابزارهای نانومتری" جمع آوری شده است[۳۲].
- ۲۲ تصویر از یک مقاله در حوزه معدن با عنوان "تخمین مقاومت برشی درزه‌های طبیعی با الگوریتم بیان ژنی" جمع آوری شده است[۳۳].
- ۱۳ تصویر از یک مقاله در حوزه پزشکی با عنوان "Galectin-3 در ایجاد فیبروز و نارسایی قلبی" جمع آوری شده است[۳۴].
- ۱۲ تصویر از یک مقاله در حوزه معدن با عنوان "ارائه یک الگوریتم چندمرحله‌ای برای شناسایی و تفکیک

[۱] Scientific Information Database (SID)



(ب) (الف)

شکل (۱): (الف) تصویری از یک صفحه سند با اطلاعات زائد (ب) تصویر پردازش شده همان سند توسط نرم افزار کازام با درجه تفکیک ۷۰۸×۶۸۲ پیکسل در پیکسل.



(ب) (الف)

شکل (۲): تصویر نمونه از مجموعه داده به همراه تصویر برچسب گذاری شده.(الف) تصویر اصلی اسناد و تصویر (ب) تصویر برچسب گذاری شده است.

زون‌های دگرسانی گرمابی در محدوده استان کرمان ماهوارهای "ASTER" جمع آوری شده است[۳۵].

- ۸ تصویر از کتاب "قصه قوهای وحشی" با شکل‌های نقاشی شده جمع آوری شده است.
- ۲ تصویر شامل جدول‌های جمع آوری شده است.

یکی از چالش‌های مهم مجموعه داده‌های غیرفارسی، کافی نبودن تعداد شکل (اطلاعات گرافیکی) درون تصویر اسناد مجموعه داده است[۳۲]. در این راستا،در مجموعه داده پیشنهادی از کتب اول و دوم ابتدایی و روزنامه‌های ابرار ورزشی، ایران ورزشی و انواع لوگوهای ایرانی که دارای اطلاعات گرافیکی بیشتری هستند استفاده شده است. از سوی دیگر، جهت جمع‌آوری اطلاعات غیرگرافیکی مانند متن و سرصفحه، از کتب دبیرستان و روزنامه‌های کیهان، آسیا، اطلاعات، ابتکار، مقاله، پایان‌نامه، کتاب دست‌نویس قدیمی منتظم ناصری و فایل های PDF فارسی استفاده شده است.

با توجه به اینکه هدف از ارائه این مجموعه داده، ناحیه‌بندی نواحی مورد نظر شامل پاراگراف، سرصفحه، لوگو، شکل، رابطه ریاضی و جدول درون تصاویر اسناد فارسی است، فقط از اسنادی که شامل چنین محتویاتی هستند استفاده شده است. به‌عنوان‌مثال، شکل(۱-الف) حاوی اطلاعات غیرضروری برای ناحیه‌بندی است (مانند نوارابزار PDF خوان و نوارابزار پس‌زمینه سیستم) که این اطلاعات زائد برای تصاویر مجموعه داده حذف شده‌اند (شکل(۱-ب)). در این تحقیق، برای انتخاب محدوده تصویر اسناد از نرم‌افزار متن‌باز کازام[۱] استفاده شده است که به صورت دستی تصویر جدید را از تصویر اسناد (مشتمل بر اطلاعات زائد) با وضوح بالا و در قالب رنگی تفکیک می‌کند. یکی دیگر از روش‌های مرسوم جهت تولید تصویر، استفاده از نرم‌افزار تبدیل PDF به عکس است که PDF را با وضوح بسیار بالایی به عکس تبدیل می‌کند. از این روش جهت تهیه تصاویر این مجموعه داده استفاده شده است.

برای برچسب زدن تصاویر مجموعه داده، از نرم‌افزار LabelImg استفاده شده است. این نرم‌افزار نواحی مورد نظر را به صورت یک مستطیل با مختصات پیکسل ابتدایی (x_0, y_0) پیکسل انتهایی (x_1, y_1) همراه با برچسب هر گروه در یک فایل XML ذخیره می‌کند. به عنوان مثال، ناحیه پاراگراف درون یک تصویر از پیکسل با مختصات (۶۰, ۵۰) تا (۱۲۰, ۱۰۰) برچسب گذاری شده است. از آنجایی که برچسب گذاری نواحی درون تصویر اسناد یک فرآیند زمان‌بر است، فایل خروجی نرم‌افزار LabelImg که حاوی اطلاعات مربوط به هر ناحیه است با استفاده از یک برنامه پیاده سازی شده پردازش و نواحی مورد نظر با رنگ‌های آبی (پاراگراف)، سبز (جدول)، مشکی (پس‌زمینه)، نارنجی (سرصفحه)، قرمز (شکل)، صورتی (لوگو) و سفید (رابطه ریاضی) برچسب گذاری می‌شود. به عنوان مثال، یک تصویر از مجموعه داده شامل نواحی پاراگراف، جدول، سرصحفه و لوگو است، فایل XML این تصویر شامل ۴ گروه با ۸ نقطه مختصات که نشان دهنده‌ی نقاط ابتدایی و انتهایی هر ناحیه است ذخیره می‌شود. در پایان با استفاده از کتابخانه Opencv یک مربع تو پر با مختصات ذخیره شده برای هر ناحیه رسم می‌شود (شکل۲).

۲-۳ معرفی گروه‌ها

۱-۲-۳ پاراگراف

برای جمع‌آوری تصاویر حاوی اطلاعات پاراگراف، از روزنامه‌های کیهان، جام‌جم، اطلاعات، آرمان ملی، کتاب درسی شیمی سال سوم دبیرستان، کتاب دست‌نویس منتظم ناصری و کتاب عجایب المخلوقات استفاده شده است. برای تنوع در انتخاب پاراگراف (مانند ابعاد، اندازه و موقعیت) از تصاویر مختلفی نمونه‌برداری شده است. برخی از پاراگراف‌های انتخاب شده درون تصاویر به صورت یک خط و برخی دیگر از چندین خط تشکیل شده‌اند.

[۱] Kazam


هر قدر دلتان می‌خواهد در گرمخانه بمانید

(الف) (ب)

شکل (۳): دو تصویر از مجموعه داده که پاراگراف‌های متفاوت را با مربع قرمز نشان می‌دهد. (الف) پاراگراف‌های یک خطی، ستونی و با فونت متفاوت (ب) پاراگراف دست‌نویس کتاب منتظم ناصری.

بزرگ‌ترین پاراگراف برچسب‌گذاری شده در این مجموعه داده در ابعاد ۶۴۸×۵۷۸ پیکسل در پیکسل و کوچک‌ترین پاراگراف با یک خط متن در ابعاد ۶×۴۵ پیکسل در پیکسل برچسب‌گذاری شده است. از سوی دیگر، پاراگراف‌های انتخاب شده درون این مجموعه داده در موقعیت‌های مختلفی نمونه‌برداری شده است (شکل۳). تعداد کل پاراگراف‌های موجود درون تصاویر مجموعه داده ۱۱۹۵۶ مورد می باشد.

۳-۲-۲ سرصفحه

سرصفحه یا عنوان یک متن، از کلمات یا جملات کوتاه با اندازه به نسبت بزرگ‌تر و با فونت (قلم) متفاوتی تشکیل شده است. برای جمع‌آوری تصاویری که حاوی اطلاعات سرصفحه هستند از روزنامه‌های همشهری، اعتماد، آفتاب یزد، جمهوری اسلامی و کتاب دست‌نویس منتظم ناصری استفاده شده است. بزرگ‌ترین سرصفحه از روزنامه همشهری با فونت B-majid_Shadow در اندازه ۱۰۸ و ۱۰۲×۶۴۲ پیکسل در پیکسل نمونه‌برداری شده است. برخی دیگر از سر صفحه‌ها با گروه‌های شکل و جدول همپوشانی دارند. از آنجایی که استخراج این‌گونه اطلاعات حائز اهمیت است، از این‌گونه سرصفحه‌ها که با گروه‌های دیگر همپوشانی دارند درون مجموعه داده نمونه‌برداری و برای هر گروه به صورت کاملاً مجزا برچسب‌گذاری شده است. به‌عنوان‌مثال در شکل (۴-پ) گروه سرصفحه با گروه شکل همپوشانی دارد و هر دو گروه به صورت جدا برچسب‌گذاری شده است.

یکی دیگر از ویژگی‌های این مجموعه داده در انتخاب سرصفحه، نمونه‌برداری از سرصفحه‌هایی است که بصورت عمودی درون اسناد ظاهر شده‌اند. به همین دلیل از چندین تصویر با سرصفحه‌های عمودی و با فونت‌های مختلف نمونه‌برداری شده است. تعداد کل سرصفحه‌های درون تصاویر مجموعه داده ۹۷۴۸ مورد می باشد.

۳-۲-۳ شکل‌ها

برای تهیه تصاویر حاوی شکل، اغلب از روزنامه‌های ابرار ورزشی، ایران ورزشی، کتب مقطع ابتدایی، کتاب قصه قوهای وحشی، ۴ مقاله مجله مرکز اطلاعات علمی و جهاد دانشگاهی و کتاب دست‌نویس عجایب المخلوقات استفاده شده است.

این منابع از نظر محتوای شکل، نسبت به سایر منابع دارای اطلاعات بیشتری هستند. تصاویر انتخاب شده در اندازه‌های مختلف نمونه‌برداری شده‌اند. کتاب عجایب المخلوقات حاوی شکل‌های قدیمی است. بزرگ‌ترین ابعاد گروه شکل درون تصویر این مجموعه داده در ابعاد ۳۵۸×۸۴۸ پیکسل در پیکسل و کوچک‌ترین آن در ابعاد ۴۱×۳۴ پیکسل در پیکسل است. شکل‌های درون تصاویر مجموعه داده در ساختارهای مختلفی قاب‌بندی می‌شوند (قاب دایره‌ایی و مستطیلی در شکل۵). تعداد شکل‌های درون تصاویر مجموعه داده ۸۵۷۲ مورد می‌باشد.

۳-۲-۴ جدول

یکی دیگر از گروه‌های معرفی شده در این مجموعه داده، گروه جدول است که با رنگ سبز برچسب‌گذاری شده است. برای جمع‌آوری تصویر حاوی اطلاعات جدول، اغلب از کتاب درس شیمی سوم دبیرستان، روزنامه‌های دنیای اقتصاد، ایران ورزشی، ابرار ورزشی و ۶ پایان‌نامه استفاده شده است.

در این مجموعه داده، انواع مختلفی از جدول بر اساس ساختار و محتوا ارائه شده است (یک سطر و یک ستون، چند سطر و چند ستون، بدون محتوا یا بامحتوا). بزرگ‌ترین جدول نمونه‌برداری شده درون تصاویر این مجموعه داده، در ابعاد ۱۰۵۴×۳۹۷ برچسب‌گذاری شده است که متعلق به روزنامه دنیای اقتصاد است. در مقابل کوچک‌ترین جدول این مجموعه داده شامل یک سطر و یک ستون است که در ابعاد ۱۶×۱۳۷ برچسب‌گذاری شده است. لازم به ذکر است که برخی از جدول

(الف) (ب) (ج)

شکل(۴): سه نمونه از تصاویر مجموعه داده با تصاویر متناظر برچسب گذاری شده که سرصفحه‌های متفاوت با مربع قرمز نشان داده می‌شود. (الف) سرصفحه‌های عمودی با فونت‌های متفاوت (ب) سرصفحه سفید رنگ با فونت Mj_Diwani Outline (ج) تصویر همپوشانی سرصفحه با گروه شکل.

استفاده شده در این مجموعه داده شامل اطلاعات غیرفارسی هستند. به عنوان نمونه می‌توان به جدول مزایده شرکت فولاد غدیر نی‌ریز اشاره کرد که در شکل (۶) آورده شده است. تعداد جدول‌های تصاویر این مجموعه داده ۷۳۴۱ مورد می باشد.

۳-۲-۵ لوگو

گروه لوگو از زیر مجموعه گروه‌های گرافیکی است که با رنگ صورتی برچسب‌گذاری شده است. از کاربردهای تشخیص لوگو در شناسایی نوع سند و یا اعتبارسنجی یک شرکت در بین مشتریان استفاده می‌شود. برای جمع‌آوری این اطلاعات از تمام لوگوهای دانشگاه‌ها، ارگان‌های دولتی، شرکت‌های خصوصی ایرانی، لوگوهای ورزشی و لوگوهای شبکه‌های اجتماعی استفاده شده است. با توجه به اینکه برخی از روزنامه‌های ورزشی ایرانی از اطلاعات و اخبار خارجی استفاده می‌کنند، در این مجموعه داده از برخی از لوگوهای ورزشی غیرایرانی نیز نمونه‌برداری شده است.

از آنجایی که گروه لوگو زیر مجموعه اطلاعات گرافیکی است و در برخی از تصاویر با گروه شکل همپوشانی دارد، امکان تشخیص اشتباه با گروه شکل وجود دارد. به همین دلیل، برای نمونه‌برداری گروه لوگو از ابعاد متفاوتی استفاده شده است. یکی دیگر از ویژگی‌های انتخاب گروه لوگو، رنگ لوگو است. برخی از لوگوها سیاه سفید و برخی دیگر به صورت رنگی نمونه‌برداری شده‌اند (شکل ۷). تعداد لوگوهای موجود درون تصاویر مجموعه داده ۷۳۲۹ مورد می باشد.

۳-۲-۶ رابطه ریاضی

گروه رابطه ریاضی از زیر مجموعه گروه‌های غیرگرافیکی و گرافیکی می باشد که با رنگ سفید برچسب‌گذاری شده است. برای جمع‌آوری این اطلاعات از ۲۳ تصویر که در مجموع شامل ۳۴۵ رابطه مشتق، انتگرال، لگاریتم، مثلثاتی، دنباله حسابی، دنباله هندسی و سایر روابط ریاضی هستند استفاده شده است. همچنین از ۵۱ تصویر دست‌نویس جزوات ریاضی مشتمل بر ۴۷۳ رابطه ریاضی نیز نمونه برداری شده است. شکل(۸) دو نمونه از تصاویر رابطه ریاضی را نشان می‌دهد.

۳-۳ داده‌افزایی مجموعه داده پیشنهادی

یکی از چالش‌های مهم شبکه‌های عصبی کانولوشنی، عدم وجود مجموعه داده کافی و کارآمد برای آموزش این شبکه‌ها است. به عبارت دیگر، در اکثر مسائل دنیای واقعی داده‌های میلیونی و برچسب‌دار برای آموزش شبکه‌های عصبی وجود ندارد. بنابراین، برخی از مسائل مطرح در حوزه درک و ناحیه‌بندی اسناد با مشکل کمبود داده [۴, ۲۰, ۵, ۲۱] و برخی دیگر با مشکل تعداد قلیل گروه‌ها مواجه هستند [۴]. در این راستا، برخی از محققان تلاش کرده‌اند که مشکل کمبود داده را با تولید داده‌های مصنوعی جبران نمایند[۵]. با توجه به اینکه جمع آوری تصویر اسناد و برچسب گذاری آن‌ها یک فرآیند زمان‌بر است، در این تحقیق، از روش‌های

(الف) (ب) (ج) (د)

شکل (۵): چهار نمونه از تصاویر مجموعه داده که شکل‌های متفاوتی را با مربع قرمز رنگ نشان می‌دهند. (الف) شکل‌های درون مقاله که حاوی اعداد و متن هستند (ب) تصویر نمونه برداری شده از روزنامه که حاوی شکل‌های زینتی هستند (ج) تصویر کتاب قصه قوهای وحشی با شکل‌های بدون قاب (د) تصویر کتاب عجایب المخلوقات با شکل‌های نقاشی شده.

(الف) (ب) (ج)

شکل (۶): سه تصاویر مجموعه داده که جدول متفاوتی را با مربع قرمز نشان می‌دهند. (الف) بزرگترین جدول نمونه‌برداری شده، (ب) کوچکترین جدول نمونه‌برداری شده است، و (ج) جدول با محتوای غیرفارسی.

(الف) (ب) (ج) (د) (ه) (و)

شکل (۷): شش تصاویر مجموعه داده که لوگوهای متفاوتی را با مربع قرمز نشان داده می‌شود. (الف) لوگو به صورت سیاه سفید، (ب) لوگو دانشگاه‌های ایران و (ج) لوگوهای غیر ایرانی (د) لوگو تیم ورزشی ملوان (ه) و (و) لوگوهای شبکه‌های اجتماعی.

[۱] Image Data Augmentation

(ب) (الف)

شکل (۸): دو نمونه از تصاویر رابطه ریاضی در مجموعه داده پیشنهادی. (الف) تصویر نوشته شده رابطه ریاضی با استفاده از نرم‌افزار (ب) تصویر دست‌نویس رابطه ریاضی.

(د) (ج) (ب) (الف)

(ح) (ز) (و) (ه)

شکل (۹): هشت تصویر مجموعه داده. (الف) تصویر روزنامه بدون نویز (ب) تصویر کتاب دست‌نویس قدیمی منتظم ناصری بدون نویز (ج) تصویر لوگو فروشگاه نادریان با نویز پواسون (د) تصویر لوگو شرکت MACK PARTS با نویز فلفل نمکی (ه) تصویر لوگو شرکت ماشین سازی تاشا با چرخش ۳۰ درجه (و) تصویر لوگو شرکت پخش پوشاک مامک با چرخش ۳۰ درجه و نویز فلفل نمکی (ز) تصویر لوگو شرکت MACK PARTS با الگوریتم برعکس‌کردن چپ به راست (ح) تصویر لوگو شرکت MACK PARTS با الگوریتم برعکس‌کردن بالا به پایین.

داده‌افزایی (مانند ایجاد نویز، چرخش[۱] و انتقال[۲] تصویر) برای افزایش تصاویر مجموعه داده پیشنهادی استفاده شده است.

از سوی دیگر، استخراج ویژگی از تصاویر همراه با نویز و یا با ساختار متفاوت از اهمیت چشمگیری برخوردار است. به همین دلیل وجود یک مجموعه داده کافی و کارآمد جهت آموزش مدل‌های یادگیری می‌تواند در بهبود فرآیند اجرا نقش بسزائی داشته باشد.

همان طور که در جدول (۲) مشاهده می‌شود مجموعه داده پیشنهادی در ابتدا مشتمل بر ۲۰۸۵ تصویر بوده است. تعداد ۱۰۸ تصویر به صورت تصادفی برای نویز پواسون انتخاب شده و به مجموعه داده افزوده شده است. به همین ترتیب از نویز فلفل نمکی روی ۱۰۸ تصویر دیگر استفاده شده است. هر یک از چرخش‌های ۳۰، ۳۳۰، ۲۷۰ و ۱۸۰ درجه روی ۱۰۸ تصویر اعمال و به مجموعه داده افزوده شده است. به همین طریق، ۱۰۸ تصویر نیز به اندازه ۳۰ پیکسل از بالا و ۳۰ پیکسل از سمت چپ به سمت پایین انتقال داده شده است. همچنین، از ۸۴۷ تصویر جهت اعمال الگوریتم‌های برعکس‌کردن (آینه‌ای کردن) چپ به راست[۳]، برعکس کردن (آینه‌ای کردن) بالا به پایین[۴] و چرخش تصاویر استفاده شده است.

[۱] Rotate
[۲] Translation
[۳] Flipping Left to Right
[۴] Flipping Up to Down

جدول (۱):داده افزایی مجموعه تصویر براساس نویز، چرخش و انتقال

تعداد تصویر	نویز	چرخش (درجه)	انتقال
۲۰۸۵	×	×	×
۱۰۸	پواسون	×	×
۱۰۸	فلفل نمکی	×	×
۱۰۸	×	۳۰	×
۱۰۸	فلفل نمکی	۳۰	×
۱۰۸	×	۲۷۰	×
۱۰۸	×	۱۸۰	×
۱۰۸	×	×	انتقال یافته
۱۰۸	پواسون	۳۰	×
۱۰۸	×	۲۷۰	انتقال یافته
۸۴۷	برعکس‌کردن چپ به راست (آینه‌ای کردن)	×	×
۸۴۷	برعکس‌کردن بالا به پایین (آینه‌ای کردن)	×	×
۸۴۷	×	۹۰	×

(الف) (ب)

شکل (۱۰): نمونه ایی از تشخیص گروه شکل توسط معیار صحت. مربع سبز رنگ نشان دهنده ی ناحیه برچسب خورده هدف و مربع قرمز رنگ نشان دهنده ی ناحیه‌بندی پیش‌بینی شده توسط مدل است. (الف) تشخیص درست ناحیه گروه شکل را توسط مدل نشان می‌دهد. (ب) تشخیص نادرست بخشی از ناحیه گروه شکل را نشان می‌دهد.

چند نمونه از تصویر مجموعه داده که بر روی آن‌ها چرخش و یا نویز اعمال شده در شکل(۹) آورده شده است. لازم به ذکر است که بر روی هیچکدام از تصاویر مجموعه داده، دو نویز یا دو چرخش اعمال نشده است.

۴ آزمایش‌ها و نتایج

۴-۱ معیارهای ارزیابی

در ادامه معیارهای ارزیابی متداول و شناخته شده‌ای که در این تحقیق استفاده شده‌اند معرفی می‌شوند.

۴-۱-۱ معیار صحت

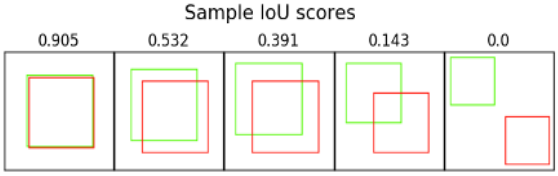
همان‌گونه که از نام این معیار مشخص است، معیار صحت[1] معرف میزان دقت شناسایی هر یک از نواحی توسط مدل است. در این راستا، برای هر یک از گروه‌ها، معیار صحت به عنوان نسبت بین تعداد پیکسل‌های درست ناحیه‌بندی شده به تعداد کل پیکسل‌های شناسایی شده برای آن گروه خاص تعریف می‌شود. به عبارت دیگر، هدف اصلی معیار صحت این است که چه سهمی از پیکسل‌های پیش‌بینی شده توسط مدل برای یک گروه خاص، درست ناحیه‌بندی شده است.

به‌عنوان‌مثال، در شکل(۱۰) این معیار برای مسائل ناحیه‌بندی تصاویر اسناد فارسی و برای گروه شکل نشان داده شده است. در این شکل، هر چقدر مربع سبز رنگ با مربع قرمز رنگ همپوشانی بیشتری داشته باشند معیار صحت بالاتری به دست می‌آید. با تجمیع کلیه گروه‌ها، در رابطه (۱) نحوه محاسبه این معیار ارائه شده است:

$$Precision = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FP_c} \qquad (۱)$$

که در آن TP_c معرف تعداد پیکسل‌های درست ناحیه‌بندی شده برای گروه c و FP_c تعداد پیکسل‌های به اشتباه ناحیه‌بندی شده برای گروه c است.

[1] Precision



شکل (۱۱): نمایش روش معیار ارزیابی IOU. مربع سبز رنگ نشان دهنده‌ی ناحیه برچسب خورده هدف و مربع قرمز رنگ نشان دهنده‌ی ناحیه پیشبینی شده از مدل است. اگر حد آستانه IOU برابر ۰/۵ قرارداد شود، دو نمونه پیش‌بینی از سمت چپ بهترین پیش‌بینی است.

۴-۱-۲ معیار فراخوان

به همراه معیار صحت، معیار فراخوان[۱] معرف قابلیت مدل در شناسایی پیکسل‌های هر گروه از نواحی است. به عبارت دیگر، از کل پیکسل‌های هر گروه چه سهمی از آن‌ها به عنوان پیکسل‌های درست آن گروه خاص شناسایی شده است. با تجمیع گروه‌ها، رابطه (۲) نحوه محاسبه معیار فراخوان را نشان می‌دهد.

$$Recall = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FN_c} \qquad (۲)$$

که در آن TP_c معرف تعداد پیکسل‌های درست ناحیه‌بندی شده برای گروه c و FN_c تعداد پیکسل‌هایی است که برای گروه c به درستی ناحیه‌بندی نشده است.

۴-۱-۳ معیار F

برای مقایسه مدل‌ها، میزان اهمیت معیارهای صحت و فراخوان بستگی به ماهیت مسئله دارد. برای داشتن یک تک معیار جهت ارزیابی مدل‌ها، معیار F[۲] از تلفیق معیارهای صحت و فراخوان معرفی شده است. این معیار، میانگین هارمونیک این دو معیار بوده و به عدد کوچکتر متمایل است. رابطه (۳) نحوه محاسبهٔ معیار **F** را نشان می‌دهد.

$$F\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (۳)$$

۴-۱-۴ معیار IOU

معیار IOU(Intersection Over Union) یک معیار ارزیابی شناخته شده است که بیشتر برای الگوریتم‌های تشخیص اشیا درون تصویر مورد استفاده قرار می‌گیرد. این معیار، شباهت ناحیه هدف و ناحیه پیش‌بینی شده را اندازه‌گیری می‌کند و عددی بین ۰ و ۱ را تولید می‌کند. این معیار هر چه به مقدار ۱ نزدیک‌تر شود نشان دهنده ناحیه‌بندی دقیق‌تر است، رابطه (۴) به همراه روش شکل(۱۱) نحوه محاسبه معیار IOU را نشان می‌دهند.

$$IOU = \frac{Area\ of\ Overlap\ region}{Area\ of\ Union\ region} \qquad (۴)$$

۵ معرفی مدل‌های کانولوشنی

در این تحقیق، برای تعیین میزان کارایی مجموعه داده پیشنهادی در کاربردهای ناحیه‌بندی مبتنی بر یادگیری عمیق، ۳ نمونه از برجسته‌ترین مدل‌های کانولوشنی انتخاب شده مورد ارزیابی قرار گرفته‌اند. در ادامه، مدل‌های انتخاب شده به اختصار معرفی شده‌اند.

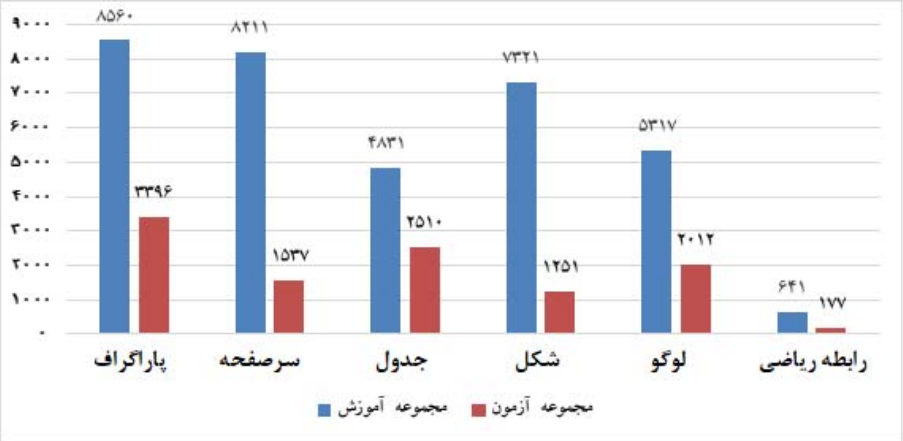
FCN-8[۳]: یکی از مدل‌هایادگیری عمیق تمام کانولوشنی است که از ۲ بخش رمزگذار و رمزگشا تشکیل شده است[۳۶]. در این مدل، از بلوک‌های کانولوشنی VGG16 برای بخش رمزگذار و برای بازسازی تصاویر از چند لایه کانولوشنی نمونه افزایی[۴] در بخش رمزگشا استفاده شده است[۳۷]. این مدل کانولوشنی مشتمل بر ۱۶ لایه کانولوشنی رمزگذار VGG16 و ۵ لایه کانولوشنی رمزگشا می‌باشد.

VGG-UNET: یک مدل تمام کانولوشنی است که برای اولین‌بار در دانشگاه فرایبورگ آلمان برای درک و ناحیه‌بندی تصاویر پزشکی ابداع شده است[۳۸]. این مدل متشکل از دو بخش رمزگذار و رمزگشا است که برای قسمت رمزگذار از بلوک‌های کانولوشنی VGG16 که قبلا توسط مجموعه داده ImageNet آموزش دیده شده است، استفاده می‌کند. مدل VGG-UNET از نظر ساختاری با مدل تمام کانولوشنی FCN-8 متفاوت است[۳۹]. این مدل از ۱۶ لایه کانولوشنی رمزگذار VGG16 و ۱۲ لایه کانولوشنی رمزگشا تشکیل شده است.

Deeplab: یکی از اهداف و چالش‌های مهم شبکه‌های عصبی کانولوشنی، استخراج نواحی کوچک و بزرگ از تصویر با کمترین تعداد پارامتر است [۴۰]. با هدف بهبود در استخراج این نواحی، روش لایه کانولوشنی فضایی معرفی شده است. شبکه یادگیری عمیق Deeplab از این ایده برای استخراج ویژگی درون تصویر استفاده می‌کند. این مدل دارای سه نوع مدل کانولوشنی

---

[۱] Recall
[۲] F-score
[۳] Fully Convolution Network
[۴] Upsampling



شکل (۱۲): توزیع گروه‌ها در مجموعه آموزش و آزمون

است. در نوع اول، از معماری کانولوشنی فضایی[۱] و میدان تصادفی شرطی کاملا متصل (CRF)[۲] برای کنترل وضوح تصویر درون این معماری استفاده می‌شود [۴۱]. نوع دوم، از ترکیب چندین فیلتر در ابعاد مختلف برای بهبود استخراج ویژگی تصویر استفاده می‌کند[۴۰]. نوع سوم، برای بهبود عملکرد نوع دوم از الگوریتم نرمال‌سازی در بین لایه‌ها استفاده می‌کند [۴۲]. در این تحقیق از Deeplab نوع سوم استفاده شده است. برای بهبود عملکرد Deeplab از مدل کانولوشنی آموزش دیده ResNet50 استفاده شده است که ۵۰ لایه کانولوشنی متعلق به مدل ResNet50 و ۱۲۲ لایه کانولوشنی دیگر برای قسمت رمزگشا استفاده می‌شود.

۶ نتایج و خروجی

در این بخش، عملکرد مدل‌های کانولوشنی بر روی مجموعه داده معرفی شده را مورد بررسی قرار می‌دهیم. محیط استفاده شده برای پیاده‌سازی مدل‌ها، گوگل کولب با GPU مدل Tesla K80 و حافظه رم ۱۲ گیگا بایت می باشد.

تصاویر مجموعه آموزش و آزمون به گونه‌ایی از یک دیگر تفکیک شده است که نسبت توزیع هر گروه در مجموعه آموزش تقریبا ۸۰% و مجموعه آزمون تقریبا ۲۰% است.

شکل (۱۲) گروه‌های مجموعه آموزش و آزمون را نشان می‌دهد که هر دو از یک توزیع نسبتاً متعادلی برخوردار هستند. به بیان دیگر توزیع گروه‌ها در مجموعه داده آموزش تقریبا ۸۰% است. نهایتا، ۴۴۷۰ تصویر برای مجموعه آموزش و ۱۱۲۸ تصویر برای مجموعه آزمون به صورت دستی جداسازی شده است.

تعداد دوره‌های[۳] آموزشی برای هر مدل کانولوشنی ۶۰ تنظیم شده است. از الگوریتم آدام[۴] با نرخ یادگیری ۰/۰۰۰۳ برای بهینه‌سازی پارامترها استفاده شده است. جدول (۳) نتایج اجرای ۳ مدل کانولوشنی بر روی مجموعه داده‌های آموزش و آزمون آورده شده است.

زمان صرف شده جهت آموزش مدل کانولوشنی VGG_UNET و Deeplab تقریباً ۱ ساعت است. از سوی دیگر، با توجه به اینکه تعداد پارامترهای مدل FCN-8 نسبت به ۲ مدل کانولوشنی دیگر بیشتر است، نیاز به زمان بیشتری جهت آموزش دارد و شاهد ضعیف‌ترین نتیجه بر روی مجموعه داده آموزش و آزمون نسبت به دو مدل کانولوشنی دیگر هستیم.

شکل (۱۳) ماتریس سردرگمی[۵] سه مدل کانولوشنی را برای ۱۱۲۸ تصویر آزمون نشان می‌دهد. یکی از ساده‌ترین راه‌حل‌های محاسبه ماتریس سردرگمی برای مسائل ناحیه‌بندی معنایی، محاسبه بر اساس تعداد پیکسل‌های درست پیش‌بینی به تعداد کل پیکسل‌های هر گروه است. در شکل(۱۳-الف) ماتریس سردرگمی مدل کانولوشنی VGG_UNET آورده شده است. با توجه به شکل شاهد هستیم که بیشترین احتمال پیش‌بینی مدل کانولوشنی برای گروه رابطه ریاضی با مقدار ۸۸%و کمترین احتمال پیش‌بینی برای گروه پس‌زمینه است. به عبارت دیگر، در این مدل ۸۸% پیکسل‌های گروه رابطه ریاضی به درستی پیش‌بینی شده است. شکل(۱۳-ب)، ماتریس سردرگمی مدل کانولوشنی FCN-8 را نشان می‌دهد. مشاهده می شود که مدل کانولوشنی FCN-8 با ۶۹ میلیون پارامتر در تشخیص پیکسل‌های گروه‌ها نسبت به ۲ مدل کانولوشنی دیگر عملکرد ضعیف‌تری ارائه نموده است. در شکل(۱۳-ج) ماتریس سردرگمی مدل کانولوشنی Deeplab ارائه کرده است. در این مدل که دارای ۱۲ میلیون پارامتر است، شاهد هستیم که در مقایسه با مدل کانولوشنی FCN-8 در شناسایی گروه‌ها عملکرد مناسبی داشته است. این بهبود عملکرد به واسطه وجود لایه‌های کانولوشنی پیچشی فضایی در بین لایه‌های مدل کانولوشنی Deeplab است.

---

[۱] Atrous Convolution

[۲] Conditional Random Field

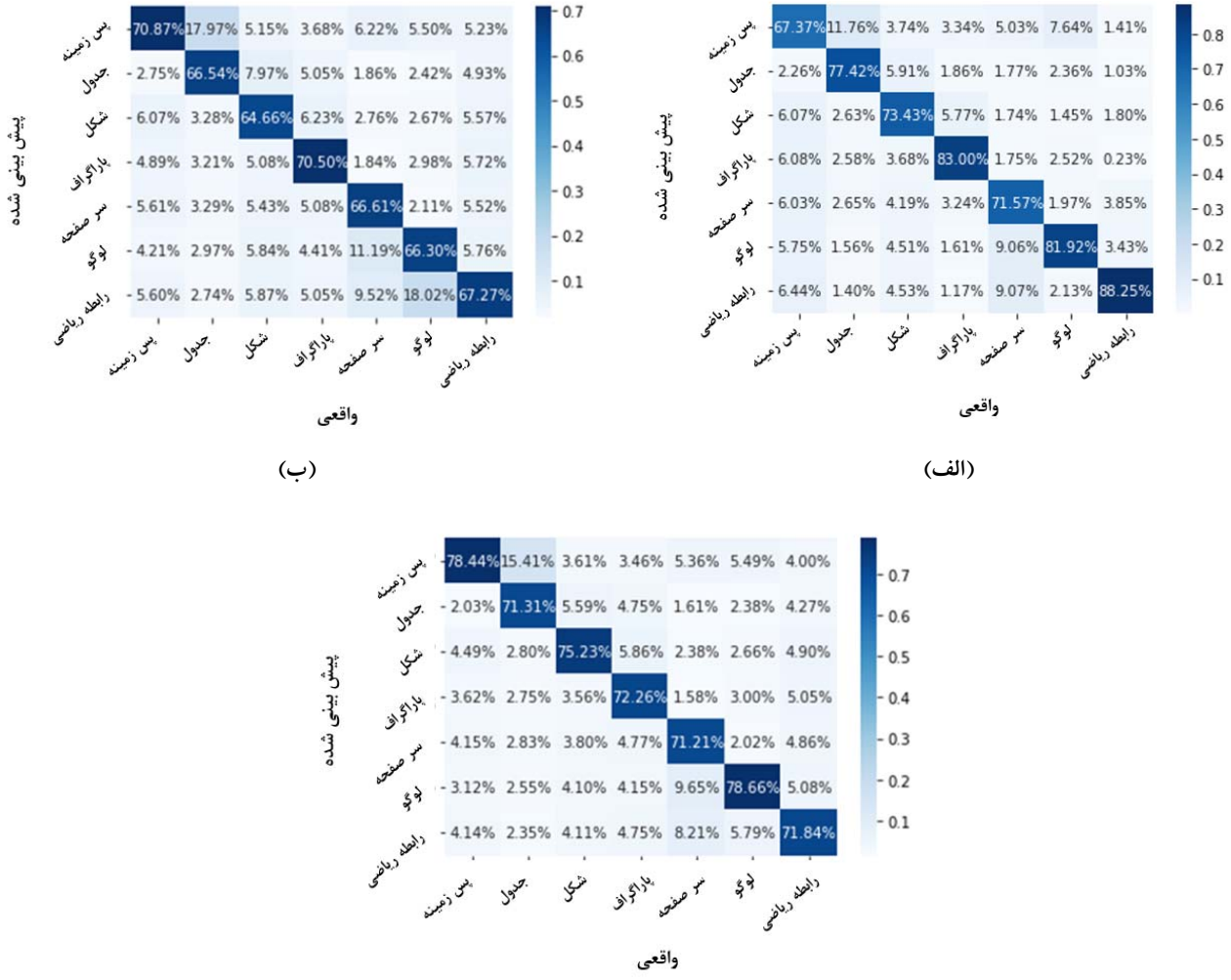
[۳] Epochs

[۴] Adam

[۵] Matrix Confusion

جدول (۲):مقایسه مدل‌های کانولوشنی روی مجموعه داده های آموزش و آزمون

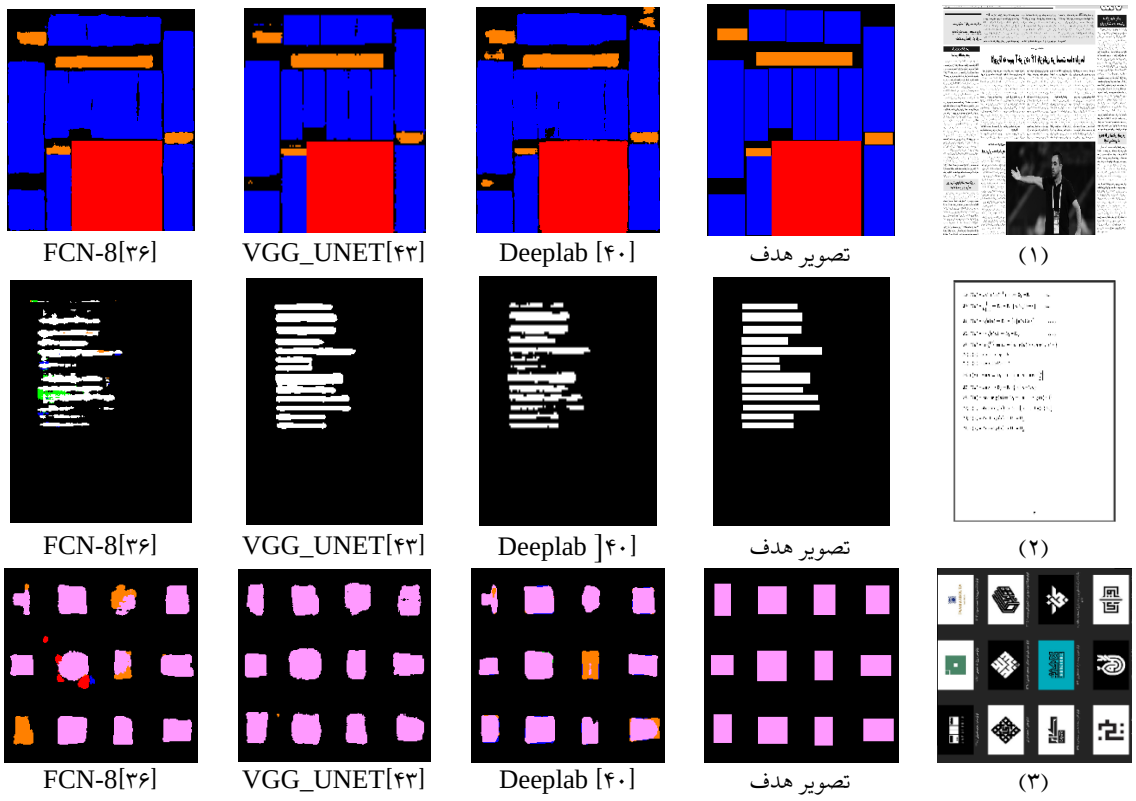
مدل	تعداد پارامترها(میلیون)	مجموعه داده	صحت	فراخوان	معیار F	IOU
VGG- UNET [۴۳]	۱۴	آموزش	۰/۷۴۹	۰/۷۵۳	۰/۷۵۷	۰/۸۴۱
		آزمون	۰/۷۲۱	۰/۷۲۲	۰/۷۲۲	۰/۸۳۵
Deeplab[۴۰]	۱۳	آموزش	۰/۷۴۱	۰/۷۴۱	۰/۷۵۲	۰/۷۶۴
		آزمون	۰/۶۸۱	۰/۶۸۱	۰/۷۵۱	۰/۷۶۴
FCN-8 [۳۶]	۶۹	آموزش	۰/۷۱۲	۰/۷۱۳	۰/۷۲۱	۰/۷۲۵
		آزمون	۰/۶۵۷	۰/۶۵۴	۰/۶۷۴	۰/۶۸۳



شکل (۱۳): ماتریس سردرگمی مرحله آزمون (الف)مدل کانولوشنیVGG_UNET[۴۳](ب) مدل کانولوشنی FCN-8 [۳۶] (ج) مدل کانولوشنی Deeplab[۴۰]

شکل(۱۴) خروجی ۳ مدل کانولوشنی را برای ۳ تصویر نمونه از مجموعه داده نشان می‌دهد.

در این شکل شاهد هستیم که مدل FCN-8 در شناسایی گروه لوگو عملکرد مناسبی نداشته است. همچنین این مدل در تشخیص



شکل (۱۴): خروجی سه مدل کانولوشنی بر روی سه تصویر نمونه

گروه‌های پس‌زمینه فی‌مابین پاراگراف‌ها عملکرد مناسبی نداشته و آنها را به عنوان بخشی از پاراگراف های مجاور شناسایی کرده است.

مدل کانولوشنی VGG-UNET با ۱۴میلیون پارامتر، خروجی مناسبی را نسبت به دو مدل کانولوشنی دیگر ارائه کرده است. یکی از دلایل تولید خروجی مناسب این مدل کانولوشنی، استفاده از الگوریتم نرمال‌سازی جهت بهینه‌سازی پارامترها و آموزش سریع مدل است. این الگوریتم، ورودی هر لایه کانولوشنی را به عددی در فاصله ۱- تا ۱ تبدیل می‌کند. مدل کانولوشنی Deeplab به دلیل استفاده از لایه پیچشی فضایی، در درک و ناحیه‌بندی گروه لوگو نسبت به مدل کانولوشنی FCN-8 بهتر عمل کرده است.

۷ جمع‌بندی

در این تحقیق،اولین مجموعه داده برای درک و ناحیه‌بندی تصاویر بر روی اسناد فارسی معرفی شده است. مجموعه داده معرفی شده شامل۵۵۹۸تصویرصفحاتروزنامهوکتبعلمی فارسی در قالب ۶گروه پاراگراف، شکل،جدول،لوگو، رابطه ریاضی و سرصفحه که به ترتیب با رنگ‌های آبی، قرمز، سبز، صورتی، سفید و نارنجی برچسب‌گذاری شده است. برای هر یک از این گروه‌ها، تجزیه و تحلیل کاملی به همراه ویژگی‌هایتصاویر آن ارائه شد. با هدف ارزیابی مناسب بودن مجموعه داده معرفی شده برای توسعه مدل‌های یادگیری عمیق، از ۳ مدل شبکه کانولوشنی شناخته شده در این حوزه استفاده شد. برای ارزیابی مدل‌های استفاده شده، از معیارهای شناخته شده صحت، فراخوان، معیار F و IOU استفاده شد و در نهایت نتایج مقایسه سه مدل کانولوشنی توسط ماتریس سردرگمی و خروجی آن‌ها ارائه شد.

مراجع

[1] A. Antonacopoulos, C. Clausner, C. Papadopoulos and S. Pletschacher., "Competition on historical book recognition," *In 12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1459-1463, 2013.

[2] A. Antonacopoulos, C. Clausner, C. Papadopoulos and S. Pletschacher., "Competition on recognition of documents with complex layouts," *In 13th International Conference on Document Analysis and Recognition (ICDAR)*, p. 1151–1155, 2015.

[3] C. Clausner, A. Antonacopoulos and S. Pletschacher, "Competition on recognition of documents with complex layouts," *In 14th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, p. 1404–1410, 2017.

[4] C. Clark and S. Divvala, "Pdffigures2.0: Mining figures from research papers," *In 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, p. 143–152, 2016.

Common objects in context," *In European conference on computer vision* (*ECCV*), p. 740–755, 2014.

[17] M. Everingham, S. M. Eslami, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision,* p. 98–136, 2015.

[18] H. Caesar, J. Uijlings and V. Ferrari, "Coco-stuff: Thing and stuff classes in context.," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 1209-1218, 2018.

[19] G. Neuhold, T. Ollmann, S. Rota Bulo and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," *In Proceedings of the International Conference on Computer Vision* (*ICCV*), *Venice, Italy,* pp. 4990-4999, 2017.

[20] X. Tao, Z. Tang, C. Xu and Y. Wang, "Logical labeling of fixed layout pdf documents using multiple contexts," *In 2014 11th IAPR International Workshop on Document Analysis Systems* (*DAS*), p. 360–364, 2014.

[21] M. T. Luong, T. D. Nguyen and M. Y. Kan, "Logical structure recovery in scholarly articles with rich document features," *Multi media Storage and Retrieval Innovations for Digital Library Systems,* pp. 270-292, 2012.

[22] M. Haurilet, Z. Al-Halah and R. Stiefelhagen, "SPaSe - Multi-Label Page Segmentation for Presentation Slides," *Conference: IEEE Winter Conference on Applications of Computer Vision* (*WACV*), pp. 726-734, 2019.

[23] K. Hadjar and R. Ingold, "Arabic Newspaper Page Segmentation," *Proceedings of the Seventh International Conference on Document Analysis and Recognition* (*ICDAR*), vol. 3, pp. 895-899, 2003.

[24] S. S. Bukhari, T. M. Breuel, A. Asi and J. El-Sana, "Layout analysis for arabic historical document images using machine learning," *in 2012 International Conference on Frontiers in Handwriting Recognition,* pp. 639-644, 2012.

[25] R. Elanwar, W. Qin, M. Betke and D. Wijaya, "Extracting text from scanned Arabic books: a large-scale benchmark dataset and a fine-tuned Faster-R-CNN model," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 24, p. 349–362, 2021.

[26] H. M. Al-Barhamtoshy, K. M. Jambi, S. M. Abdou and M. Rashwan, "Arabic Documents Information Retrieval for Printed, Handwritten, and Calligraphy Image," *IEEE Access,* vol. 9, pp. 51242-51257, 2021.

[27] M. Potanin, D. Dimitrov, A. Shonenkov, V. Bataev, D. Karachev, M. Novopoltsev and A. Chertok, "Digital Peter: New dataset, competition and handwriting recognition methods," *Computer vision and pattern recognition,* pp. 43-48, 2021.

[28] I. Rabaev, B. K. Barakat, A. Churkin and J. El-Sana,

[5] X. Yang, E. Yumer, P. Asente, M. Kraley, D. Kifer and C. Lee Giles, "Learning to extract semantic structure from documents using multi modal fully convolutional neural networks.," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 5315-5324, 2017.

[6] L. Gao, X. Yi, Z. Jiang, L. Hao and Z. Tang, "Competition on page object detection," *In 14th IAPR International Conference on Document Analysis and Recognition* (*ICDAR*), vol. 1, pp. 1417-1422, 2017.

[7] S. A. A. A. Arani, E. Kabir and R. Ebrahimpour, "Handwritten Farsi word recognition using NN-based fusion of HMM classifiers with different types of features," *International Journal of Image and Graphics,* vol. 19, no. 01, 2019.

[8] M. Cordts, O. Mohamed, R. Sebastian, E. Timo, E. Markus, B. Rodrigo, F. Uwe, R. Stefan and S. Bernt, "The cityscapes dataset for semantic urban scene understanding," *In Proceedings of the IEEE conference on computer vision and pattern recognition* (*CVPR*), pp. 3213-3223, 2016.

[9] L. G. Hafemann, R. Sabourin and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," *Pattern Recognition,* vol. 70, pp. 163-176, 2017.

[10] R. Smith, "An overview of the Tesseract OCR engine," *In Proceedings of the Ninth International Conference on Document Analysis and Recognition* (*ICDAR 2007*), vol. 2, p. 629–633, 2007.

[11] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2015.

[12] G. Lin, C. Shen, A. Van Den Hengel and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), p. 3194–3203, 2016.

[13] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *In International Conference on Learning Representations* (*ICLR*), 2014.

[14] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," *arXiv preprint arXiv:1503.02351,* 2015.

[15] Z. Zhang, A. G. Schwing, S. Fidler and R. Urtasun, "Monocular object instance segmentation and depth ordering with cnns," *arXiv preprint arXiv:1505.03159,* pp. 2614-2622, 2015.

[16] T. Y. Lin, M. Michael, B. Serge, H. James, P. Pietro, R. Deva, D. Piotr and C. L. Zitnick, "Microsoft coco:

"The HHD Dataset," *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2020.

[29] S. Shirali-Shahreza, M. T. Manzuri-Shalmani and M. H. Shirali-Shahreza, "Page segmentation of Persian/Arabic printed text using ink spread effect," *In 2006 SICE-ICASE International Joint Conference,* pp. 259-261, 2006.

[30] A. A. Shirazi, A. Dehghani, H. Farsi and M. Yazdi, "Persian Logo Recognition using Local Binary Patterns," *3rd International Conference on Pattern Recognition and Image Analysis (IPRIA),* pp. 258-261, 20147.

[31] A. U. Islam, M. J. Khan, M. Asad, H. A. Khan and K. Khurshid, "iVision HHID: Handwritten hyperspectral images dataset for benchmarking hyperspectral imaging-based document forensic analysis," *Data in Brief,* vol. 41, p. 107964, 2022.

[۳۲] ا. چراغی, ن. بحرانی و ر. ملک‌فر, "بررسی تأثیر نانوتکنولوژی بر علوم پزشکی و زیست محیطی از دیدگاه ابزارهای نانومتری," *حیات*, ۹۴-۸۵۱۳۸۳.

[۳۳] م. شمس الدین سعید, س. کریمی نسب و ح. جلالی فر, "تخمین مقاومت برشی درزه های طبیعی با الگوریتم بیان ژنی," *نشریه مهندسی معدن*, ۸۷-۷۶, ۱۴۰۱ .

[۳۴] ر. عطابخشیان, ف. رایگان و ف. کازرونی, "نقش Galectin-3 در ایجاد فیبروز و نارسایی قلبی," *مجله تعالی بالینی*، سال دوم شماره ۲ (پیاپی ۴، تابستان ۱۳۹۳), ۳۶-۴۹, ۱۳۹۳.

[۳۵] م. عراق نسترن, ا. اصغری, س. مجیدی فر و س. طالش حسینی, "ارائه یک الگوریتم چندمرحله‌ای برای شناسایی و تفکیک زون‌های دگرسانی گرمابی در محدوده استان کرمان ماهواره ای ASTER," *انجمن مهندسی معدن ایران*, ۲۸-۳۹, ۱۴۰۱ .

[36] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), p. 3431–3440, 2015.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[38] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Computer Vision and Pattern Recognition, pp. 234-241, 2015.

[39] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," IEEE conference on computer vision and pattern recognition, pp. 248-255, 2009.

[40] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, p. 834–848, 2018.

[41] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," arXiv:1412.7062, 2014.

[42] L. C. Chen, G. Papandreou, F. Schroff and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," arXiv preprint arXiv:1706.05587, 2017.

[43] V. Iglovikov and A. Shvets, "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation," Computer Vision and Pattern Recognition, 2018.

[44] J. Bhatt, K. A. Hashmi, M. Z. Afzal and D. Stricker, "A Survey of Graphical Page Object Detection with Deep Neural Networks," Applied Sciences, p. 5344, 2021.

امین فرجی دانش آموخته کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی در دانشگاه شهید باهنر کرمان است. زمینه‌های پژوهشی ایشان یادگیری‌ماشین، پردازش‌تصویر و بینایی‌ماشین است.

مسعود سعید در سال ۱۳۷۴ مدرک کارشناسی خود را در رشته ریاضی-کاربرد در کامپیوتر، از دانشگاه صنعتی شریف دریافت کرد. سپس مدرک کارشناسی ارشد و دکتری خود را در رشته مهندسی کامپیوتر به ترتیب از دانشگاه علم و صنعت و شیراز در سال‌های ۱۳۷۸ و ۱۳۹۶ دریافت نمود. در حال حاضر، ایشان استادیار بخش مهندسی کامپیوتر دانشگاه شهید باهنر کرمان است. زمینه‌های تحقیقاتی مورد علاقه ایشان سیستم های توصیه گر،داده کاوی و شبکه های عصبی است.

حسین نظام آبادی پور در سال ۱۳۷۷ مدرک کارشناسی خود را در رشته مهندسی برق- الکترونیک از دانشگاه شهید باهنر کرمان دریافت کرد. سپس، مدرک کارشناسی ارشد و دکتری را در رشته مهندسی برق-الکترونیک از دانشگاه تربیت مدرس، به ترتیب در سال های ۱۳۷۹ و ۱۳۸۳ دریافت کرد. ایشان در سال ۱۳۸۳ به عنوان استادیار به بخش مهندسی برق دانشگاه شهید باهنر کرمان پیوست و در سال ۱۳۹۱ به درجه استادی ارتقا یافت. دکتر نظام آبادی پور نویسنده و هم-نویسنده بیش از ۴۰۰ مقاله در ژورنال‌ها و کنفرانس‌های علمی بوده است. زمینه‌های علاقه‌مندی ایشان، شامل پردازش‌تصویر، بازشناسی‌الگو، رایانش‌نرم و الگوریتم‌های فراابتکاری است.