مقاله مروری

http://jmvip.sinaweb.net
www.ismvip.ir

این مقاله در قالب نهایی آن در مجله
ماشین‌بینایی و پردازش‌تصویر چاپ خواهد شد
تاریخ انتشار برخط: خردادماه ۱۴۰۰

مروری بر روش‌های یادگیری ژرف در بازشناسی نوری نویسه‌ها با تاکید بر رسم الخط‌های فارسی، عربی و اردو

شیما کاشف[۱]، حسین نظام‌آبادی‌پور[۲] و الهام شعبانی‌نیا[۳]

چکیده

در سال‌های اخیر موفقیت شبکه‌های عصبی کانولوشنی ژرف در تشخیص و بازشناسی اشیا سبب جلب توجه بسیاری از حوزه‌های مختلف یادگیری ماشین، از جمله حوزه بازشناسی نوری نویسه‌ها، به این مقوله شده است. یکی از چالش‌های مهم این حوزه، استخراج ویژگی‌های متمایزکننده و حاوی اطلاعات است. غالب روش‌هایی که در سال های گذشته در بازشناسی نوری حروف مطرح شدند، مبتنی بر ویژگی‌های دستی هستند که توانایی تعمیم‌پذیری محدودی دارند. امروزه به کمک شبکه‌های کانولوشنی می‌توان استخراج ویژگی را به صورت خودکار و با کارایی فوق العاده‌ای به ماشین سپرد و بازشناسی حروف جدا، ارقام و لیگاتورها را بادقت بالایی انجام داد. همچنین، ساختارهایی بر مبنای ترکیب شبکه های کانولوشنی و بازگشتی پیشنهاد شده اند، که می توانند بدون نیاز به جداسازی حروف، بازشناسی را انجام دهند. این رویکرد در سال‌های اخیر مورد توجه زیاد محققان بینایی ماشین قرار گرفته است؛ چرا که به کمک این شبکه‌ها می‌توان به شکل مستقل از زبان، بازشناسی را تنها با توجه به مجموعه آموزشی انجام داد. هدف از این مقاله، مروری بر کارهای انجام شده با این رویکرد نوین در حوزه بازشناسی نوری نویسه‌ها است. در ادامه، پس از بیان مسئله و مروری مختصر بر روش‌های قبل، روش‌های مبتنی بر الگوریتم های ژرف و ویژگی‌های آن‌ها با تفصیل بیشتری ارزیابی می‌شوند. از آنجا که تاکید این مقاله روی تحقیقات بازشناسی نوری حروف در رسم الخط های پیوسته، نظیر فارسی، عربی و اردو است، کارهای انجام شده در این حوزه ها نیز در بخشی جداگانه مرور می شوند. همچنین، ضمن معرفی مجموعه‌های داده معروف برای کاربردهای مختلف و مروری بر معیارهای ارزیابی روش های بازشناسی نوری حروف، مهم ترین نرم افزارهای اختصاصی و بسته های نرم افزاری متن بازی که برای بازشناسی حروف استفاده می‌شوند، معرفی خواهند شد.

کلیدواژه‌ها

بازشناسی نوری نویسه ها، الگوریتم های یادگیری ژرف، پایگاه‌های داده، رسم الخط فارسی، عربی و اردو

این مقاله در آبان‌ماه ۱۳۹۹ دریافت در فروردین‌ماه ۱۴۰۰ بازنگری و پذیرفته شد.

[۱] گروه پژوهشی سیستم‌های هوشمند واژه، کرمان، ایران،
رایانامه: s.kashef@vajeh-research.org
[۲] بخش مهندسی برق، دانشگاه شهید باهنر کرمان، کرمان، ایران،
رایانامه: nezam@uk.ac.ir
[۳] دانشکده مهندسی کامپیوتر، دانشگاه صنعتی سیرجان، سیرجان، ایران،
رایانامه: eshabaninia@sirjantech.ac.ir

مولف مسئول: شیما کاشف

۱. مقدمه

متن مجموعه ای از نمادها است که برای ضبط، برقراری ارتباط و انتقال فرهنگ استفاده می شود. متن به عنوان یکی از تأثیرگذارترین اختراعات بشر، نقش مهمی در زندگی وی داشته است. OCR مخفف عبارت Optical Character Recognition به معنای نویسه خوان نوری است و وظیفه ی آن تشخیص خودکار متن ها در تصاویر و اسناد و تبدیل آن به

متن قابل جستجو و ویرایش در رایانه است. یک تصویر یا یک سند ممکن است از نظر انسان ارزش اطلاعاتی بسیاری داشته باشد اما از دید کامپیوتر، آن سند فقط متشکل از چند پیکسل ساده در یک تصویر است. برای این که بتوانیم از اطلاعات نوشتاری در تصاویر در رایانه استفاده کنیم باید از نرم افزارهای OCR کمک بگیریم. نرم افزار نویسه خوان نوری، متن اسناد را می خواند و آن را به قالب قابل ویرایش در رایانه تبدیل می کند. با این که تصاویر پویش شده در رایانه بسیار حجیم هستند و امکان جستجو در آن ها وجود ندارد، خروجی نرم افزار های نویسه خوان، بسیار کم حجم هستند و می توان به راحتی یک متن را در آن جستجو کرد. نرم افزارهای مبتنی بر OCR در کاربردهای وسیعی استفاده می شوند [۱]. از جمله این کاربردها می توان به بازشناسی دستنوشته ها [۲]، تشخیص خودکار پلاک خودرو [۳, ۴]، خواندن چک های بانکی و اعتبار سنجی امضای افراد [۵]، کپچا[1] [۶]، کتابخانه های دیجیتال و بازشناسی نوری نت‌های موسیقی اشاره کرد. همچنین، اطلاعات معنایی غنی و دقیق درون متن در طیف گسترده ای از کاربردهای مبتنی بر تصویر، مانند جستجوی تصویر [۷]، بازرسی هوشمند [۸]، اتوماسیون صنعتی [۹] و مسیریابی با ربات ها [۱۰] اهمیت دارد.

با وجود اینکه تحقیق در زمینه OCR سابقه طولانی دارد، به دلیل وجود چالش های مختلف، پژوهش های علمی همچنان در این حوزه در حال انجام هستند. به طور مثال، متن‌های چند زبانه یا با اندازه‌های مختلف قلم، پرسپکتیو و نورپردازی نامناسب در تصویر برداری با دوربین، تصاویر تار به واسطه حرکت دوربین، پس زمینه شلوغ یا تاریک تصاویر، برخی از این چالش ها هستند. در مورد رسم الخط فارسی، عربی و اردو، پیوستگی حروف و شباهت زیاد برخی از حروف به یکدیگر (برخی از حروف، تنها در مکان یک نقطه با یکدیگر تفاوت دارند) نیز به این چالش ها می افزاید و کار را برای تشخیص متون سخت می کند. همچنین، به واسطه تعداد کم‌تر استفاده کنندگان این زبان ها نسبت به زبان هایی نظیر انگلیسی و چینی، تعداد پژوهش ها در این زبا نها کم تر است.

نگارش فارسی، عربی و اردو، ویژگی های منحصر به فردی دارد که آن را کاملا از نگارش لاتین متمایز می سازد. در ادامه، ویژگی های کلی رسم الخط فارسی مرور می شود که در مورد دو رسم الخط عربی و اردو نیز صادق است [۱۱]. همچنین، شکل ۱، این ویژگی ها را به طور خلاصه نشان می دهد.

- متون فارسی بر خلاف متون لاتین از راست به چپ نوشته می شوند.
- در کلمات فارسی، برخی از حروف از یک یا دو طرف به حروف مجاور خود اتصال دارند و برخی نیز به صورت مجزا نوشته می شوند. در نتیجه هر کلمه ممکن است شامل یک یا چند بخش متصل باشد که "زیر کلمه" نامیده می شود (شکل ۱-الف).
- حروف فارسی به صورت متصل به هم نوشته می شوند و کلمات را تشکیل می دهند. این امر، بازشناسی متون فارسی را نسبت به متون لاتین دشوارتر می سازد.
- حروف فارسی ممکن است چهار شکل مجزا با توجه به محل قرار گیری در کلمه داشته باشند (شکل ۱-ب).

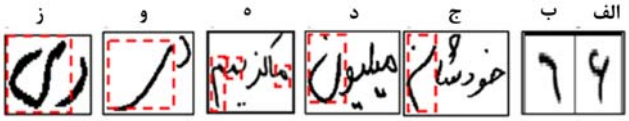
شکل۱. الف) چسبیده بودن حروف به طور کلی و مجزا بودن تعدادی از حروف در نگارش فارسی، ب) چهار شکل مختلف حرف "ع" با توجه به موقعیت آن در کلمه، ج) هم پوشانی دو حرف "ح" و "ک" در کلمه "حکم"، د) اتصال حروف "ک" و "ا" در دو محل، ه) حروف متفاوت با بدنه مشابه، و) کشیدگی حرف "ب" در کلمه "با" [۱۱].

- حروف واقع در یک کلمه ممکن است همپوشانی داشته باشند؛ بدین معنا که نتوان با یک خط عمودی، حروف را به طور کامل از یکدیگر مجزا کرد (شکل ۱-ج).
- برخی از حروف شکل یکسانی دارند و تفاوت آن ها تنها در تعداد نقطه یا جایگاه نقطه ها در آن حروف است. همچنین دو حرف "ک" و "گ" تنها در وجود یه سرکش با یکدیگر تفاوت دارند (شکل ۱-ه).
- حروف فارسی ممکن است در بالا یا پایین بدنه، دارای اعراب، تشدید، علامت همزه یا تنوین باشند. هرچند کاربرد اعراب در زبان فارسی نسبت به زبان عربی محدودتر است، اما اگر کلمه ای نامتداول باشد یا به دلیل تشابه نگارشی آن به کلمه دیگر، تأکید بر تلفظ صحیح آن باشد، از نشانه های اعراب استفاده می شود.
- حروفی که از طرف چپ قابلیت اتصال به حرف مجاور خود را دارند، ممکن است به صورت کشیده نوشته شوند (شکل ۱-و).
- بیشتر حروف فارسی (مخصوصا حروف چسبیده) دندانه دار هستند. در مواردی که کیفیت سند اصلی یا دستگاه اسکنر پایین باشد، ارتفاع دندانه ها نسبت به خط طمینه کوتاه می

[1] Captcha

شود و این امر، شناسایی صحیح این حروف را با مشکل مواجه می کند.

در مورد خط دستنویس فارسی، مشکلات بازشناسی متون، به مراتب بیشتر می شود که از جمله آن ها می توان به کشیده نوشتن برخی حروف و شکل متفاوت حروف یکسان اشاره کرد. برخی از این موارد در شکل ۲ آورده شده است.



شکل ۲. برخی از چالش های بازشناسی خط دستنویس فارسی: الف وب) دو شکل نوشتاری عدد ۶، ج و د) حرف "ن" در جایگاه یکسان (پایان کلمه) با شکل نوشتاری متفاوت، ه) شکل متفاوت نگارش م در در کلمه "ماکزیمم"، و و د) کشیده نویسی حرف "ی" در کلمه "دی" [۱۲].

به نظر می رسد، مرجع [۱۳] اولین مرجع برای بازشناسی حروف چاپی فارسی باشد. مرجع [۱۴] یک سامانه بازشناسی نوری حروف فارسی یکپارچه را مبتنی بر معماری تخته سیاه پیشنهاد داده است که در آن تلاش شده است انواع عوامل موثر بر بازشناسی خط فارسی، نظیر ویژگی های آماری، ماژول تشخیص قلم، ماتریس درهم ریختگی، جهت نوشتن، اطلاعات مربوط به نقطه ها و علائم و استفاده از لغت نامه واژه ها در نظر گرفته شود.

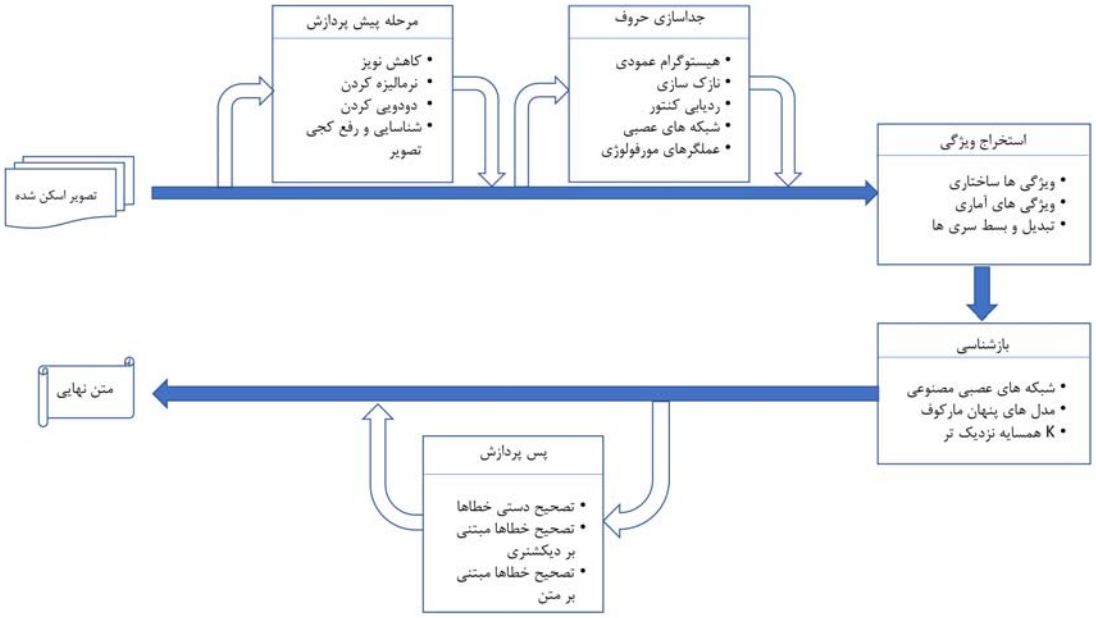
روش‌های بازشناسی نوری حروف را می توان از منظرهای مختلف نظیر چاپی [۱۴] یا دستنویس بودن [۱۵]، پیوستگی حروف (پیوسته/گسسته)، دامنه مجموعه لغات (محدود/نامحدود) و نوع نگارش (مقید/آزاد) دسته بندی کرد. همچنین، تقسیم بندی‌های دیگری برای متون دستنویس لحاظ شده است. دسته بندی اول، دسته بندی برون خط-برخط است که به ترتیب به اسناد پویش شده و دستنوشته هایی که از طریق قلم نوری و یک صفحه رقومی کننده[۲] ثبت می شوند، اشاره می کنند. در روش های بر خط، علاوه بر اطلاعات مربوط به موقعیت قلم، اطلاعات زمانی مربوط به مسیر قلم نیز در اختیار است. در این روش می توان از اطلاعات زمانی سرعت، شتاب، فشار و زمان برداشتن و گذاشتن قلم روی صفحه در بازشناسی استفاده کرد [۱۶]. این سیستم در رایانه های دستی[۳] و همچنین در برخی از آخرین نسخه های گوشی های لمسی موجود می باشد [۱۷]. در بازشناسی برون خط، از تصویر دو بعدی متن ورودی استفاده می شود. در این روش به هیچ نوع وسیله نگارش خاصی نیاز نیست و تفسیر داده ها مستقل از فرایند تولید آن ها و تنها بر اساس تصویر متن صورت می گیرد. این روش به نحوه بازشناسی به وسیله انسان شباهت بیشتری دارد [۱۱]. از کاربردهای بازشناسی برون خط، می توان به خواندن آدرس های پستی [۱۸]، چک ها [۱۹] و فرم ها اشاره کرد. دسته بندی دیگر اسناد، به پیوسته یا گسسته نوشتن حروف برمی-گردد. به طور مثال، برای پرکردن برخی فرم های خاص نظیر کنکور، نیاز است حروف نام و نام خانوادگی به صورت گسسته درون کادرهای مشخص، ثبت شوند. این در حالی است که خط فارسی، یک خط پیوسته است [۲۰].

در این مقاله، روش های بازشناسی نوری حروف را از منظر روش های مستقل از الگوریتم های ژرف و روش های مبتنی بر الگوریتم های ژرف، تقسیم کرده ایم. روش های مبتنی بر الگوریتم های ژرف، با ظهور شبکه های ژرف پدید آمدند. روش های مستقل از الگوریتم های ژرف، خود به دو دسته روش های مبتنی بر جداسازی[۴] و روش های بدون نیاز به جدا سازی[۵] تقسیم می شوند. در روش های مبتنی بر جدا سازی، پس از یافتن متن در تصویر، حروف تشکیل دهنده کلمات از یکدیگر جدا می شوند و به سامانه شناسایی داده می شود. واحد شناسایی، یک طبقه بند است که قبلا با حروف مشابه آموزش دیده است. بدیهی است که در متونی که درجه تفکیک[۶] پایینی داشته باشند و همچنین در زبان هایی که حروف به طور پیوسته نوشته می شوند، اگر جداسازی حروف به درستی انجام نشود دقت سامانه را به شدت پایین می آورد. نقص دیگری که این روش ها دارند این است که برای شناسایی حروف، تنها به تصویر بریده شده از همان حرف اکتفا می کنند و اطلاعاتی از حروف قبل و بعد از آن ندارند. حال آنکه به کمک این اطلاعات، بازشناسی متن دقت بالاتری می تواند داشته باشد [۲۱]. در روش های مستقل از جداسازی، نیازی به جداسازی حروف از یکدیگر نیست و این روش ها می توانند به طور مستقیم روی کلمات و خطوط شامل کلمات کار کنند. از سوی دیگر، روش های مبتنی بر الگوریتم های ژرف، غالبا مبتنی بر جداسازی نیستند. البته این بدان معنی نیست که کاملا بدون نیاز به جداسازی باشند؛ بلکه در حد جداسازی خطوط یا واژه ها، به جداسازی نیاز دارند. در ادامه، تاریخچه و ساختار کلی روش های مستقل از الگوریتم های ژرف را به طور خلاصه مرور می کنیم.

[2] Digitizer
[3] Handheld PC-PDA

[4] Segmentation-based methods
[5] Segmentation-free methods
[6] Resolution

یک سامانه OCR روش های مستقل از الگوریتم های ژرف، از تعدادی اجزاء تشکیل شده، که برای مثال در شکل ۳ نشان داده شده است. همانطور که در این شکل مشخص است، برخی از مراحل شامل مراحل پیش پردازش، جداسازی حروف و پس پردازش، اختیاری هستند و با توجه به روش مورد نظر، می توانند در طول فرآیند استفاده شوند. مرحله اول، پویش اسناد است، تا تصویر اسناد به صورت دیجیتالی در دسترس باشد. در کاربردهای OCR در منظره، تصاویر حاوی متن، با دوربین های دیجیتال گرفته می شود.



شکل ۳. نمونه ای از بازشناسی متن در روش های مستقل از الگوریتم های ژرف

در مرحله بعد، انواع روش های پیش پردازش برای بهبود کیفیت تصویر انجام می شود. هدف این مرحله، افزایش خوانایی تصویر متن و حذف جزئیاتی از تصویر است که هیچ توانایی تمایز دهی در فرآیند بازشناسی ندارند [۲۲, ۲۳]. از مهم ترین این روش ها، می توان به دودویی کردن[7]، کاهش نویز [۲۴]، ارتقا تصویر [۲۵]، نرمالیزه کردن [۲۶]، شناسایی و رفع کجی[8] [۲۷] و چرخش[9] تصویر [۲۸] اشاره کرد. همچنین، مرجع [۲۹]، روش پیش پردازشی را مختص رسم الخط های فارسی و عربی، برای شناسایی و جبران سازی ناپیوستگی های نا مطلوب در زیر کلمه ها پیشنهاد داده است.

استخراج ویژگی ها در مرحله بعد قرار دارد. هدف از استخراج ویژگی، بدست آوردن خصوصیات و ویژگی های اساسی تصویر است. این ویژگی ها که در مرحله بازشناسی استفاده قرار می شوند، باید به گونه ای باشند که حروف یا کلمات را به خوبی از هم تمایز دهند [۳۰]. روش های استخراج ویژگی، از یک کاربرد تا کاربرد دیگر متفاوت هستند. روش هایی که در یک کاربرد موفق هستند، ممکن است در کاربردهای دیگر به خوبی عمل نکنند [۳۱]. بنابراین، انتخاب روش مناسب برای استخراج ویژگی، یکی از مهم ترین مراحل برای دستیابی به دقت بالا در بازشناسی است. روش های استخراج ویژگی به سه دسته کلی تقسیم می شوند؛ الف) ویژگی های ساختاری، ب) ویژگی های آماری و ج) تبدیل و بسط سری ها[10] [۳۲]. ویژگی های آماری با شمارش ویژگی های محلی در هر پیکسل و استخراج مجموعه ای از آمار از توزیع ویژگی های محلی، توزیع مکانی پیکسل ها را تجزیه و تحلیل می کنند. مهم ترین ویژگی آماری حروف، ناحیه بندی[11] است که حروف به نواحی هم پوشان و غیر هم پوشان تقسیم می شوند و توزیع چگالی پیکسل

[7] Binarization
[8] Slant
[9] Skew
[10] Transformation and series expansions
[11] Zoning

های حروف در نواحی مختلف، تجزیه و تحلیل می شود [۳۳-۳۵]. ویژگی های ساختاری رایج ترین ویژگی های مورد استفاده محققان است. این ویژگی ها مشخصات هندسی و توپولوژیکی تصویر متن را با توصیف خصوصیات محلی و فرامحلی آن‌ها نشان می دهند. ویژگی های ساختاری به دسته الگویی که باید طبقه بندی شود، بستگی دارد. به طور مثال، برای متن های فارسی و عربی، ویژگی ها شامل نقاط و جایگاه آن ها، سرکش ها و طول و عرض آن ها، جهات و تقاطع بخش ها و حلقه ها هستند [۳۶, ۳۷]. روش های تبدیل فرامحلی[12]، توانایی تغییر نمایش پیکسل ها به شکل فشرده تری را دارند. این روش ها می توانند سیگنال را با ترکیب خطی از یک سری توابع ساده تر نشان دهند. از تبدیل های رایجی که برای بازشناسی حروف استفاده می شود، می توان به تبدیل فوریه [۳۸]، تبدیل کسینوسی گسسته [۳۹]، ویولت ها [۴۰]، تبدیل هاف [۴۱] و گشتاورها [۴۲] اشاره کرد.

در نهایت، مرحله بازشناسی است که این کار، با توجه به ویژگی های استخراج شده هر یک از حروف جداسازی شده انجام می شود. روش های رایج در طبقه بندی عبارتند از شبکه های عصبی مصنوعی [۴۳-۴۵]، مدل های پنهان مارکوف [۴۶-۴۸] و k همسایه نزدیک تر [۴۹].

روش های بازشناسی، از دو رویکرد مبتنی بر جداسازی کلمات به حروف و زیر حروف و رویکرد مبتنی بر شکل کلی کلمات استفاده می کنند [۵۰]. جداسازی حروف، در روش های مبتنی بر جداسازی، یک مرحله بسیار مهم است؛ زیرا میزان دقت این مرحله به طور مستقیم روی نرخ بازشناسی صحیح تاثیر می گذارد [۵۱]. این مرحله، به خصوص در زبان های پیوسته، نظیر فارسی و عربی از اهمیت ویژه ای برخوردار است. روش های مختلفی برای جداسازی وجود دارد. یکی از این روش ها، استفاده از هیستوگرام افقی است [۱۳, ۴۲, ۵۲]. در حالیکه از هیستوگرام عمودی برای جداسازی کلمات، زیر کلمات و حروف استفاده می شود، هیستوگرام افقی برای جدا سازی خطوط به کار می رود. جداسازی بر اساس نازک سازی[13] [۵۳]، روش های مبتنی بر ردیابی کانتور [۵۴-۵۶]، روش های مبتنی بر شبکه های عصبی مصنوعی [۵۷] و روش های مبتنی بر عملگرهای مورفولوژی [۵۸] از دیگر روش های جدا سازی پر کاربرد هستند.

روش های بازشناسی با رویکرد دوم، بر مبنای تشخیص شکل کلی کلمه ها و زیر کلمه ها کار می کنند [۵۹]. همچنین، در مواردی که درجه تفکیک اسناد پویش شده پایین است و جداسازی حروف به خوبی امکان پذیر نیست، روش های مبتنی بر شکل کلی جایگزین بهتری هستند [۶۰]. از جمله مراجعی که از این رویکرد برای بازشناسی مستقیم کلمه ها از طریق زیر کلمه های تشکیل دهنده آن ها استفاده کرده اند، می توان به تحقیق [۶۱] اشاره کرد. در مرجع [۶۲]، لغت نامه ای از ۱۱۳،۳۴۰ زیر کلمه چاپی، با ۴ نوع قلم و ۳ اندازه متفاوت گردآوری شده است. این زیر کلمه ها به وسیله خوشه بند k-میانگین، به ۳۰۰ خوشه مجزا تقسیم شده اند. برای بازشناسی کلمه جدید، ابتدا تعیین می شود هر زیر کلمه به کدام خوشه تعلق دارد. بعد از پیدا کردن شکل کلی کلمه ها، استفاده های متفاوتی از آن ها می تواند صورت بگیرد. ممکن است از شکل کلی برای جداسازی دقیق تر حروف استفاده شود یا می توان با نقاط و علامت ها، زیر کلمه را تشخیص داد. همچنین، می توان شکل کلی را برای تایید زیر کلمه بازشناسی شده با جداسازی، به کار گرفت [۶۳]. البته، روش های مبتنی بر شکل کلی، ایرادهایی هم دارند که مزایا و معایب آن ها به طور کامل در مرجع [۶۴] آمده است. یکی از مهم ترین این مشکلات، زیاد بودن تعداد کلاس ها است. در مراجعی نظیر [۶۵, ۶۶]، تحقیقاتی برای کم کردن تعداد این کلاس ها انجام شده است.

اولین روش های بازشناسی مبتنی بر شکل کلی کلمات، از مدل های پنهان مارکوف[14] (HMM) استفاده می کردند [۶۷, ۶۸]. مدل های پنهان مارکوف، روش هایی موفق در پیش بینی دنباله سری های زمانی بودند. ایده اصلی برای استفاده از HMM ها، تبدیل تصویر دو بعدی ورودی (تصویر دودویی شده) به یک بردار ویژگی یک بعدی با لغزاندن یک پنجره روی تصویر است. در اجرای روش پنجره لغزان، جزئیات متنوعی از جمله اندازه و شکل پنجره و هم پوشانی هنگام حرکت، وجود دارد [۶۹]. همچنین، ویژگی های مختلفی از تصویر با لغزاندن این پنجره روی تصویر، استخراج می شود. ویژگی های چگالی نظیر تعداد پیکسل های سیاه در پنجره [۴۸, ۷۰]، ویژگی های مربوط به تقعر [۷۱, ۷۲]، کانتور [۷۳] و راستا [۷۴] از مشهورترین این ویژگی ها هستند که در مدل های مخفی مارکوف استفاده می شوند. پس از آموزش مدل پنهان مارکوف با تصاویر برچسب خورده، برای بازشناسی تصاویر آزمون، کافی است با استفاده از پنجره لغزان، ویژگی های مشابه از تصاویر آزمون استخراج شده و به مدل آموزش دیده برای بازشناسی داده شود. از جمله کارهای انجام شده با استفاده از مدل های پنهان مارکوف در زبان های با خط پیوسته، می توان به مراجع [۲۴, ۲۶, ۷۵-۸۰] اشاره کرد. به طور مثال، مرجع [۲۴]، روشی را برای بازشناسی کلمه های دستنویس فارسی و عربی با استفاده از مدل

[12] Global transformation
[13] Thinning
[14] Hidden Markov Models

پنهان مارکوف پیشنهاد داده است. در این مقاله، هیستوگرام جهت های کد زنجیره ای[15] مربوط به نوارهای عمودی روی تصویر کلمه که با استفاده از یک پنجره لغزان از راست به چپ روبش می شود، به عنوان بردار ویژگی استفاده می شود. بردارهای ویژگی استخراج شده به عنوان ورودی به کمی ساز بردار خود سازمان ده کوهنن[16] داده می شود. اطلاعات همسایگی که در نقشه ویژگی خود سازمان ده (SOFM) نگهداری می شود، برای هموار سازی توزیع احتمال مشاهدات در HMMهای آموزش دیده استفاده می شود. دقت این روش، روی مجموعه داده ای با ۱۷،۰۰۰ تصویر از اسامی دستنویس ۱۹۸ شهر مختلف ایران به حدود ۶۵% رسید. با توجه به پیوستگی حروف در خط فارسی، دقت این روش با دقت روش های ارائه شده در آن زمان روی متون لاتین برابری می کند.

پس از بازشناسی حروف، روش های پس پردازش برای افزایش دقت بازشناسی استفاده می شوند. به طور کلی، روش های تصحیح خطا در بازشناسی حروف به سه دسته روش های دستی، مبتنی بر فرهنگ لغت و مبتنی بر متن تقسیم می شوند [۸۱, ۸۲]. به طور شهودی، راحت ترین روش برای تصحیح خطاهای بازشناسی، استفاده از نیروی انسانی است. با این حال، این روش علاوه بر اینکه کاری پر زحمت، پر هزینه و وقت گیر است، روشی همراه با خطای زیاد است.

در روش مبتنی بر فرهنگ لغت، از یک واژه نامه برای بررسی هجی کلمات بازشناسی شده، استفاده می شود و در صورت وجود غلط املایی، آن ها را تصحیح می کند. در برخی موارد، تعدادی کلمه نامزد برای اصلاح غلط املایی پیشنهاد می شود. به عبارت دیگر، یک الگوریتم انطباق رشته حروف وجود دارد که با استفاده از یک معیار فاصله، به کلمات موجود در متن، وزن هایی نسبت می دهد. کلمه نامزدی از دیکشنری که کمترین فاصله را با کلمه دارای غلط املایی داشته باشد، جایگزین آن کلمه در متن می شود. با وجودی که روش های اصلاح مبتنی بر فرهنگ لغت، به ظاهر ساده و موثر می آیند، اما مشکلات و محدودیت های آن ها، باعث شده است که این روش ها، بهترین روش های اصلاح غلط های املایی پس از بازشناسی حروف نباشند. اولین محدودیت این است که این روش ها، نیاز به یک فرهنگ لغت جامعی دارند که تمام دایره لغات موجود در یک زبان را شامل شود. در زبان فارسی، فرهنگ های لغات متعددی از جمله دهخدا، معین و عمید وجود دارد که هر سه آن ها، مشکل عدم به روز رسانی لغات را دارند؛کلماتی که در سال های اخیر وارد زبان شده اند یا واژه های منسوخی که باید از فرهنگ لغات حذف شوند. به علاوه کلمه هایی نظیر "گوگل"، که در بسیاری از متن های امروزی رایج است، اصلا در فرهنگ لغات فارسی نمی گنجد. از سوی دیگر، هرچه تعداد کلمات موجود در فرهنگ لغت بیشتر باشد، به همان اندازه دقت و سرعت سیستم های تشخیص کلمه کاهش می یابد. بعضی از سیستم ها برای کاهش مشکلات ناشی از فرهنگ لغت بزرگ، در ابتدا تعداد کلمات مورد بررسی برای تشخیص کلمه آزمون را کاهش می دهند. این کار نه تنها سرعت سیستم را بالا می برد بلکه دقت را نیز افزایش می دهد. به طور مثال، مرجع [۸۳]، تصاویر کلمات دستنوشته فارسی را بر اساس شکل کلی کلمات خوشه بندی کرده و یک فرهنگ لغت تصویری تولید کرده است. بدین صورت، یک لیست از کلمات کاندید برای تشخیص کلمه آزمون ورودی تولید می شود. این مرحله به عنوان مرحله ای برای کاهش دامنه جستجو استفاده می شود. در مرحله دوم، بهترین کاندید از لیست به دست آمده از مرحله اول برای شناسایی کلمه ورودی انتخاب می شود. مشکل دیگری که روش های پس پردازش مبتنی بر فرهنگ لغات دارند این است که یک فرهنگ لغت معمول، از یک زبان پشتیبانی می کند. بنابراین، در متن های چند زبانه نمی توان از این روش ها استفاده کرد. مسأله دیگر، این است که اکثر فرهنگ های لغت، اسامی شهرها، کشورها و افراد مشهور را در خود ندارند. همچنین در فرهنگ لغت دهخدا که این اسامی جمع آوری شده است، عدم به روز رسانی آن مشکل ایجاد می کند. آخرین محدودیت روش های مبتنی بر فرهنگ لغت این است که حتی در صورت به روز رسانی فرهنگ های لغت، آن ها ماهیتی ایستا دارند و نمی توانند هم گام با رشد وسیع واژگان سرازیر شده به زبان های مختلف، رشد کنند.

روش های اصلاح اشتباهات مبتنی بر متن، بر پایه مدلسازی آماری و n-gram های کلمات پایه ریزی شده اند [۸۴]. به صورت کلی، این مدل ها بر اساس احتمال رخداد یک مورد (حرف، کلمه، پاره گفتار) پس از دنباله ای از $n-1$ مورد عمل می کند. در زبان لاتین که حروف به صورت جدا از هم نوشته می شوند، n-gram را می توان روی حروف مجاور در کلمه ها بررسی کرد. به طور مثال، احتمال قرار گیری حرف k بعد از حرف h در زبان انگلیسی صفر است یا احتمال توالی دو حرف 'an' بیشتر از 'oi' است. در رسم الخط های پیوسته نظیر فارسی و عربی، n-gram های کلمات و پاره گفتار استفاده می شوند. به طور مثال، در جمله "درخشش تور خورشید"، با توجه به متن می توان متوجه شد که کلمه "نور" به اشتباه "تور" بازشناسی شده است. این درحالی است که روش های مبتنی بر فرهنگ لغت نمی توانند این اشتباهات را تصحیح کنند؛ زیرا هر دو کلمه "نور" و "تور" دارای معنی هستند. مرجع [۸۵]، روشی

[15] Chain-code
[16] Kohonen self-organizing vector quantization

برای بازشناسی کلمات برخط فارسی ارائه می دهد که با استفاده از بافت جمله سعی در بهبود بازشناسی دارد. در این مرجع، ابتدا علائم و بدنه زیر کلمات دست نوشته ورودی تفکیک شده و بدنه هر زیر کلمه و علائم آن مشخص می شود. سپس، علائم زیر کلمات، تشخیص داده شده و بر اساس آن، مجموعه ای از واژگان به عنوان فرضیه در نظر گرفته می شوند. به هر فرضیه بر اساس میزان شباهت آن به دستنوشته ورودی امتیاز تعلق می گیرد و بر اساس امتیاز حاصل، محتمل ترین فرضیات مشخص می شوند. سپس، این رویه به وسیله مدل زبانی، برای یافتن فرضیات محتمل تر هدایت می شود. مرجع [۸۶]، با استفاده از شاخه علمی پردازش زبان طبیعی، یک الگوریتم سه مرحله ای برای بازشناسی متون فارسی بر مبنای بازشناسی جملات فارسی ارائه می دهد. این الگوریتم در مرحله اول از یکی از روش های برچسب زنی اجزای متصل، به منظور جداسازی زیر کلمات استفاده می کند. در مرحله دوم، زیر کلمات استخراج شده کنار هم قرار می گیرند و همه کلمات معنی دار و سپس جملات بالقوه با معنی تشکیل می شوند. در آخرین مرحله، از مدل های زبانی بایگرام و ترایگرام و همچنین از قواعد گرامری زبان فارسی به منظور تشخیص جمله صحیح از میان مجموعه ای از جملات استفاده می شود.

در نهایت، به متن استخراج شده از تصویر می رسیم که آخرین مرحله از این سامانه است.

مقاله های مروری مختلفی روی انواع تحقیقات انجام شده در حوزه بازشناسی نوری حروف با استفاده از روش های مستقل از الگوریتم های ژرف صورت گرفته است [۳۱, ۳۲, ۸۷-۸۹]. با این وجود و با توجه به جدید بودن روش های مبتنی بر الگوریتم های ژرف در بازشناسی نوری حروف، تحقیقات اندکی در مرور این روش ها انجام شده است. یکی از کامل ترین این تحقیقات، مرجع [۹۰] است که کار خود را به مرور کارهای انجام شده در بازشناسی متن در منظره محدود کرده است. در این مقاله، ضمن مرور کوتاهی بر ساختار کلی روش های مستقل از الگوریتم های ژرف برای بازشناسی حروف، تمرکز اصلی، روی بررسی روش های بازشناسی حروف با استفاده از الگوریتم های ژرف است. در این راستا، مباحث و مفاهیم مورد نیاز در این زمینه به همراه پژوهش های مطرح انجام شده مرور می شوند. همچنین، مجموعه های داده استاندارد و پرکاربرد موجود در زبان های مختلف با محوریت زبان فارسی و عربی معرفی شده اند. در پایان، سامانه های تجاری/تحقیقاتی آزمایش شده در بازشناسی نوری حروف، معرفی می شوند.

۲. روش های مبتنی بر الگوریتم های ژرف در بازشناسی نوری حروف

با ظهور و پیشرفت شبکه های ژرف، استفاده از این شبکه ها در مسأله بازشناسی نوری حروف، مورد توجه پژوهشگران قرار گرفت. شبکه های بازگشتی، غیر از اینکه نوع نمایش ورودی و خروجی به چه صورت باید باشد، نیاز به هیچ دانش دیگری از داده ندارند [۹۱]. نکته بعد اینکه، شبکه های بازگشتی بر خلاف مدل های پنهان مارکوف که تصمیم گیری را تنها بر مبنای حالت پنهان فعلی انجام می دهد، اطلاعات متنی شامل حالت های قبلی را با ساختار LSTM ترکیب می کنند [۹۲]. از سوی دیگر، این شبکه ها می توانند به صورت تشخیصی آموزش ببینند و ساختار داخلی آن ها، ساز و کار بسیار قدرتمندی را برای مدلسازی سری های زمانی، فراهم می کند. همچنین، این شبکه ها در برابر نویزهای زمانی و مکانی، مقاوم هستند. نکته حائز اهمیت دیگر در مورد روش های بازشناسی حروف مبتنی بر شبکه های ژرف که در سال های اخیر مطرح شده اند، این است که بر خلاف روش های مستقل از الگوریتم های ژرف بازشناسی حروف، در روش های مبتنی بر الگوریتم های ژرف اخیر نیازی به برچسب گذاری صریح برای هر بردار ستونی از دنباله ورودی نیست. همچنین، یکی دیگر از مزایایی که سامانه های بازشناسی حروف مبتنی بر الگوریتم های ژرف دارند، قابلیت تعمیم پذیری بالاتر آن ها نسبت به روش های مستقل از الگوریتم های ژرف، برای متن های دیده نشده است. مرجع [۹۳]، برای اثبات قابلیت تعمیم پذیری روش پیشنهادی خود که بر پایه شبکه LSTM است، آموزش شبکه را با استفاده از داده های آموزشی مصنوعی انجام داد، اما کارایی شبکه را روی داده های آزمون واقعی، که شامل صفحات اسکن شده کتاب های مختلف بود، سنجید.

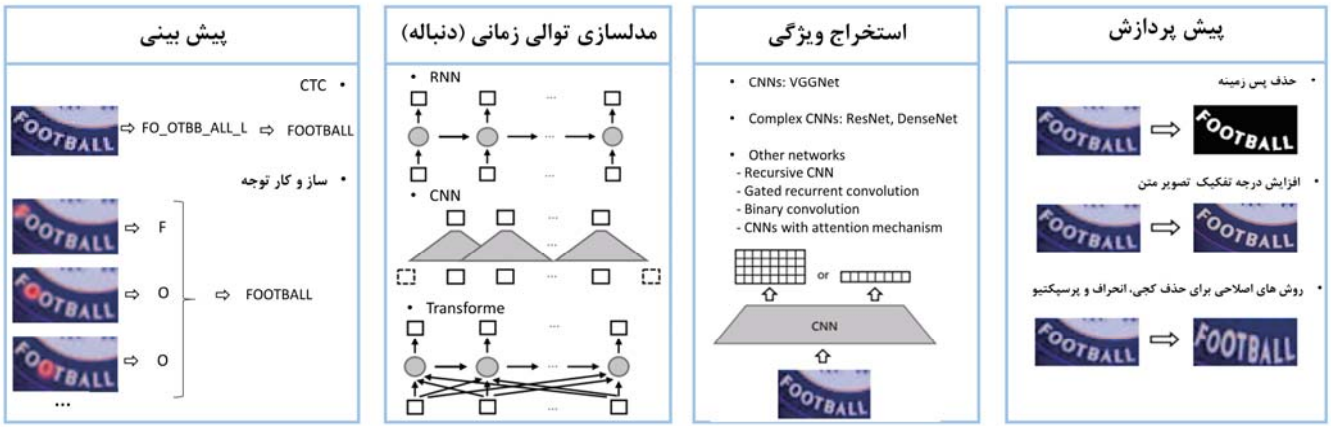
مشابه روش های مستقل از الگوریتم های ژرف، روش های مبتنی بر الگوریتم های ژرف را نیز می توان به روش های مبتنی بر جداسازی و مستقل از جدا سازی تقسیم بندی کرد. البته، اکثر قریب به اتفاق روش های مبتنی بر الگوریتم های ژرف که در سال های اخیر مطرح شده اند، مستقل از جداسازی حروف هستند؛ لیکن پژوهش های ابتدایی در استفاده از شبکه های ژرف، مبتنی بر جداسازی حروف بودند. از جمله روش های مبتنی بر الگوریتم های ژرف مبتنی بر جداسازی می توان به نسخه متن باز کتابخانه tesseract نسخه ۳ به پایین [۹۴] و اپلیکیشن PhotoOCR [۹۵] که برای استخراج و بازشناسی متن از تصاویر در گوشی های هوشمند طراحی شده است اشاره کرد. تحقیق ارائه شده در مرجع [۹۶] نیز مبتنی بر جداسازی حروف است که حروف پس از جداسازی به وسیله CNN ای

که قبلا توسط حروف برچسب خورده آموزش دیده است، شناسایی می شوند.

در معرفی روش های مبتنی بر الگوریتم های ژرف مستقل از جداسازی، ابتدا، پژوهش هایی برای استفاده از شبکه های کانولوشنی برای بازشناسی صورت گرفت، که به واسطه دقت پایین، زیاد مورد توجه محققین قرار نگرفت [۹۷]. یکی از مشکلاتی که برای استفاده از شبکه های عصبی کانولوشنی در مسأله بازشناسی نوری حروف وجود دارد، اندازه تصاویر ورودی است که بسته به اندازه کلمه مورد نظر می تواند کوتاه (در کلمه دو حرفی مثل OK) یا بلند (در کلمه ۱۵ حرفی مثل Congratulations) باشد. تلاش هایی برای حل این مشکل پیشنهاد شد. به طور مثال، مرجع [۹۸]، با مسأله بازشناسی متن در منظره به صورت یک مسأله طبقه بندی تصویر برخورد کرد، که وظیفه آن طبقه بندی 90k کلمه موجود در دیکشنری انگلیسی بود. پیچیدگی محاسباتی بالا و عدم تعمیم پذیری این روش برای بازشناسی سایر زبان ها، رغبت محققین برای ادامه این مسیر را از بین برد. OCR اساسا یک مسأله پیش بینی دنباله مبتنی بر تصویر است. در حالیکه شبکه های بازگشتی (RNN[۱۷])، مناسب ترین شبکه ها برای مسأله بازشناسی دنباله هستند، شبکه های کانولوشنی (CNN[۱۸])، مناسب ترین گزینه برای مسائل مبتنی بر تصویر هستند. بنابراین ترکیب این دو شبکه، گزینه مناسبی برای مسأله بازشناسی نوری حروف است.

چارچوب کلی در روش های مبتنی بر الگوریتم های ژرف، شناخت خط یا کلمه به عنوان یک کل و تمرکز روی نگاشت کل تصویر متن به یک دنباله از حروف و ارقام است. این کار به طور مستقیم با استفاده از یک چارچوب رمزگذار-رمزگشا[۱۹] انجام می شود و بنابراین نیازی به جداسازی حروف نیست. شکل ۴، ساختار معمول روش های مبتنی بر الگوریتم های ژرف را نشان می دهد که شامل چهار مرحله، پیش پردازش تصویر، استخراج ویژگی ها (ویژگی های مکانی)، مدلسازی توالی زمانی (دنباله) و بازشناسی است [۹۰]. به طور خلاصه، تصویر ورودی شامل متن، به یک دنباله از تصاویر کوچکتر، تقسیم می شود. ویژگی های مکانی و زمانی این وصله های تصویر[۲۰] به ترتیب با استفاده از شبکه های کانولوشنی و بازگشتی استخراج می شوند. در نهایت، یک مرحله هم تراز سازی برای پیش بینی و تولید دنباله ای از نویسه های خروجی با استفاده از دنباله حاصل از مراحل قبل، انجام می شود.

[17] Recurrent Neural Network
[18] Convolutional Neural Network
[19] Encoder-decoder
[20] Image patches



شکل ۴. ساختار روش های بازشناسی حروف مبتنی بر الگوریتم های ژرف [۹۰]

۱-۲ پیش پردازش تصویر

هدف از مرحله پیش پردازش، افزایش کیفیت تصاویر است که می تواند باعث بهبود ویژگی های استخراج شده و بازشناسی شود. روش های مبتنی بر الگوریتم های ژرف، می توانند به طور مستقیم با تصاویر رنگی و خاکستری نیز کار کنند و از این رو، دودویی کردن تصویر، لزوما از اجزاء مرحله پیش پردازش محسوب نمی شود. در بازشناسی تصاویر موجود در منظره، به واسطه پس زمینه شلوغ، تاری تصویر به دلیل حرکت دوربین و سایر عوامل، مرحله پیش پردازش تصویر اهمیت زیادی دارد. از موارد پیش پردازشی که در این تصاویر به کار برده می شود، می توان به حذف پس زمینه، افزایش درجه تفکیک تصویر و روش های اصلاحی[۲۱] نظیر حذف کجی، انحراف و پرسپکتیو متن های موجود در تصویر اشاره کرد. به طور مثال، مرجع [۹۹] با استفاده از شبکه های مولد متخاصم[۲۲] اقدام به حذف پس زمینه از تصاویر متنی موجود در منظره کرده است. در مورد روش های بهبود درجه تفکیک تصویر، روش های قدیمی نظیر دوخطی[۲۳] و دو مکعبی[۲۴] پاسخ گوی تصاویر متنی که تار شده اند، نیستند. مرجع [۱۰۰] شبکه ای طراحی کرده اند که به طور همزمان، اقدام به بهبود درجه کیفیت تصویر و بازشناسی متن می کند. به عبارت دیگر، این روش تمرکز اصلی را روی افزایش درجه کیفیت متن موجود در تصویر می گذارد، نه پس زمینه. کارهای زیادی در زمینه اصلاح انحراف موجود در متن تصاویر منظره انجام شده است. از جمله این کارها می توان به مراجع [۱۰۱-۱۰۷] اشاره کرد. در دستنوشته ها و همچنین، در متون چاپی نیز رفع انحراف موجود در خطوط، یکی از مراحل پیش پردازش است. در دستنوشته ها، انحراف خطوط امری طبیعی است و مراجعی از جمله [۱۰۸] اقدام به رفع این انحراف ها کرده اند. انحراف های موجود در متن های چاپی، می توانند به واسطه افتادن فاصله بین صفحه و اسکنر به وجود آید که می تواند منجر به دقت پایین بازشناسی حروف شود. مرجع [۱۰۹] روشی را برای حذف کجی خطوط در متن های چاپی پیشنهاد داده است که در مرجع [۹۳]، این روش بهبود داده شده است. در برخی از زبان ها نظیر انگلیسی، خط پایه[۲۵] به راحتی و با استفاده از فرا فکنی افقی[۲۶] قابل استخراج است؛ اما در کلمه های دستنویس فارسی و عربی، این روش به دلیل طرح های نوشتار متعدد نظیر شکسته نویسی یا انحنا دادن به برخی از حروف و نیز به دلیل نقطه های متعدد در کلمه ها، مناسب نیست. مرجع [۱۱۰]، یک روش دو مرحله ای برای تخمین و تصحیح خط پایه زیر کلمه های فارسی و عربی را پیشنهاد داده است. در این مقاله، بر اساس الگوریتم انطباق الگو، پیکسل های خط پایه کاندید شناسایی می شوند. مسیر نوشتار و خط پایه زیر کلمه ها به ترتیب در مرحله اول و دوم این روش، تخمین زده می شوند. پس از تخمین در هر مرحله، خط پایه در مرحله تصحیح، تنظیم می شود.

برای ساختن مدل های یادگیری عمیق مفید، خطای اعتبار سنجی[۲۷] باید با خطای آموزش، کاهش داشته باشد. اگر چنین اتفاقی نیفتد، اصطلاحا مدل، دچار بیش برازش[۲۸] شده است.

[21] Rectification methods
[22] Generative Adversarial Networks (GANs)
[23] Bilinear
[24] Bicubic
[25] Baseline
[26] Horizontal projection
[27] Validation
[28] Overfit

بیش برازش می تواند به دلیل تعداد و تنوع کم نمونه های آموزش باشد؛ بدین صورت که یادگیری شبکه تنها نسبت به نمونه های آموزش صورت می گیرد و توانایی تعمیم آن را برای نمونه های اعتبارسنجی و آزمون ندارد. برای غلبه بر این مشکل، روش هایی شامل افزایش داده ها[29]، رها کردن تعدادی از نورون ها[30]، نرمالیزه کردن دسته ای[31]، یادگیری انتقالی[32]، پیش آموزش[33]، یادگیری one-shot و یادگیری zero-shot وجود دارند. به جز روش اول که سعی در افزایش داده های آموزش با اعمال تغییراتی روی داده های آموزش موجود دارد، سایر روش ها که به روش های regularization معروف هستند، تلاش می کنند با وجود تعداد کم نمونه های آموزشی، قدرت یادگیری شبکه را به داده هایی که تا کنون ندیده است، بسط دهند [۱۱۱].

مرجع [۱۱۲]، دسته بندی جامعی روی انواع روش های افزایش داده دارد. این مرجع، روش های افزایش داده را به چند گروه اصلی شامل تبدیلات هندسی، تبدیلات فضای رنگی، فیلترهای کرنل، مخلوط کردن تصاویر، پاک کردن تصادفی، افزایش فضای ویژگی، یادگیری تخاصمی[34]، افزایش مبتنی بر شبکه های GAN[35]، انتقال سبک عصبی[36] و طرح های فرایادگیری[37] تقسیم می کند. هر یک از این روش ها، خود شامل زیر روش هایی هستند، که ورود به جزئیات آن ها، خارج از حوصله این مقاله است. به طور مثال، روش های مبتنی بر تبدیلات هندسی، شامل flipping، چرخش، تزریق نویز، جا به جایی تصاویر به چپ، راست، بالا و پایین و برش[38] است.

۲-۲ استخراج ویژگی ها

مرحله استخراج ویژگی ها، تصویر متن ورودی را به نمایشی نگاشت می کند که ویژگی های مرتبط برای بازشناسی حروف را در خود دارد، در حالیکه ویژگی های نامرتبط نظیر نوع قلم، رنگ، اندازه و پس زمینه تضعیف می شوند. به طور مثال، مرجع [۱۰۸] مجموعه ای از ویژگی های هندسی متون دستنوشته و مرجع [۹۲] ویژگی های HOG تصویر کلمات را استخراج کرده اند. توانایی شبکه های کانولوشنی در استخراج ویژگی های نهان شده در تصاویر ورودی، سبب پایداری در برابر اعوجاج های محلی موجود در تصاویر می شود. از این رو، در تحقیقات زیادی، استخراج ویژگی های مهم تصاویر متنی بر عهده این شبکه ها گذاشته شد [۹۳, ۱۱۳, ۱۱۴]. تصاویر ورودی که از روی آن ها، متن مورد نظر باید بازشناسی شود، از چند لایه کانولوشنی عبور می کنند. این لایه ها، ویژگی های مهم این تصاویر را استخراج می کنند. شبکه عصبی کانولوشنی، به ازای هر گام زمانی[39] از تصویر ورودی، یک سری ویژگی برای آن گام زمانی بدست می آورد.

۲-۳ مدلسازی توالی زمانی (دنباله)

مدل سازی توالی، به عنوان پلی بین ویژگی های بصری و بازشناسی، می تواند اطلاعات متنی را در یک دنباله از نویسه ها برای مرحله بعدی برای بازشناسی هر نویسه بگیرد. این کار، نسبت به در نظر گرفتن هر نویسه به طور جداگانه، مفیدتر و موثرتر است.

در میان شبکه های ژرف، شبکه های بازگشتی (RNNs) مناسب ترین شبکه ها برای پیش بینی دنباله هستند. با این حال، شبکه های بازگشتی، دو نقص بزرگ به نام مشکل محو شدگی و انفجار گرادیان[40] دارند. شبکه های بازگشتی به صورت یک زنجیره هستند و خروجی یکی، ورودی دیگری است. بنابراین، اگر گرادیان بزرگتر از ۱ باشد، به مرور زمان، بزرگ و بزرگ تر می شود تا به اصطلاح منجر شود. در حالت دیگر، اگر گرادیان خیلی کوچک باشد، به مرور کوچک تر می شود تا کاملا محو شود. به عبارت دیگر، مشکل محو شدگی گرادیان، ناشی از دستیابی محدود به اطلاعات پیشین دنباله در معماری RNN است که باعث می شود تاثیر ورودی های قبل روی لایه های پنهان، به صورت نمایی کاهش یابد. در نتیجه، شبکه RNN نمی تواند اطلاعات زیادی از گذشته سامانه را به حافظه بسپارد [۱۱۵]. به واسطه این مشکل، شبکه های RNN سنتی نتوانستند کارایی خوبی را در حل مسائل با مقیاس بزرگ نظیر OCR و بازشناسی گفتار از خود نشان دهند. ظهور شبکه های عمیق LSTM[41] [۱۱۶] انقلابی در شبکه های RNN بود که منجر به حل مشکلات و محدودیت های شبکه های بازگشتی، از جمله مشکل محو شدگی گرادیان در این شبکه ها شد. LSTM یک شبکه بازگشتی به شدت غیر خطی است که دارای گیت های ضرب شونده و فیدبک های جمع شونده است. شکل، ساختار یک نورون LSTM را نشان می دهد. این نورون

[29] Data augmentation
[30] Dropout
[31] Batch normalization
[32] Transfer learning
[33] Pretraining
[34] Adversarial training
[35] GAN-based augmentation
[36] Neural style transfer
[37] Meta-learning schemes
[38] Cropping

[39] Time-step
[40] Vanishing and exploding gradient problems
[41] Long short term memory

دارای سه گیت فراموشی، ورودی، و خروجی است. به طور کلی، گیت فراموشی تصمیم می گیرد چه مواردی از حالت قبل باید نگه داری شوند. گیت ورودی در مورد اینکه چه مواردی از حالت کنونی اضافه شود، تصمیم می گیرد. در نهایت، گیت خروجی تعیین می کند که حالت مخفی بعدی باید چه باشد.

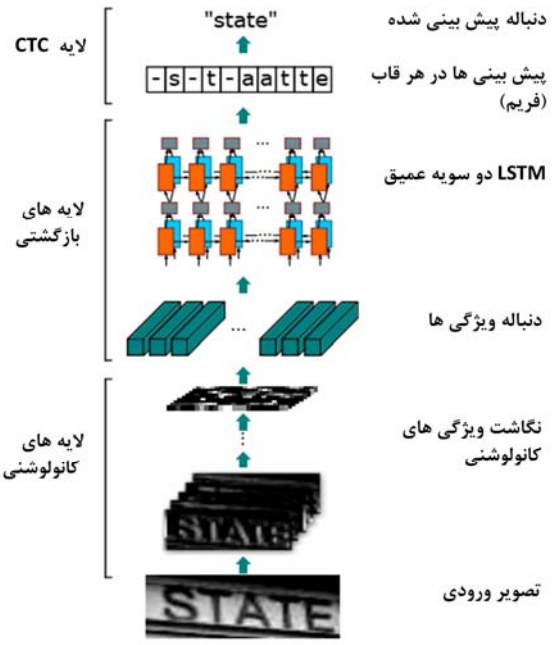
معرفی شبکه های بازگشتی چند بعدی[42] [۱۱۷] توجه برخی از محققین را به خود جلب کرد. این شبکه ها، با فراهم کردن ارتباطات بازگشتی در همه جهت ها، شبکه های بازگشتی استاندارد را توسعه دادند. در سال ۲۰۰۸، Grave و همکارانش نسخه دوسویه LSTM[43] را برای دسترسی به متن در هر دو جهت رو به جلو و عقب معرفی کردند [۱۰۸]. در بسیاری از کاربردها، دسترسی به محتوای آینده می تواند همانند دسترسی به محتوای گذشته مفید باشد. در بازشناسی حروف، به طور مثال، شناسایی یك کاراکتر می تواند هم با کمك کاراکترهای پیشین و هم با کمك کاراکترهای بعد از آن صورت گیرد. از این رو، در تحقیقات زیادی، برای لایه های بازگشتی، از شبکه های LSTM دو سویه استفاده می کنند.

۲-۴ پیش بینی و هم تراز سازی

هدف از این مرحله، تخمین دنباله نویسه ها از ویژگی های استخراج شده از تصویر متن ورودی است. در حقیقت، این مسأله یک مسأله دنباله به دنباله[44] است. همانطور که از نام این مدل ها مشخص است، هدف این مدل ها، تولید یک دنباله خروجی با توجه به دنباله ورودی است که عموما طول متفاوتی دارند. نکته مهم در مسأله های بازشناسی حروف با استفاده از روش های دنباله به دنباله، تشخیص نویسه صحیح در هر گام زمانی است، بدون اینکه هیچ دانش قبلی درباره تراز بین پیکسل های تصویر و نویسه های هدف وجود داشته باشد. به طور کلی، روش هایی که برای حل این مسأله در سال های اخیر پیشنهاد شده اند، روش های طبقه بندی زمانی پیوندی (CTC[45]) و روش های مبتنی بر ساز و کار توجه[46] هستند [۱۱۸]. اخیرا، تحقیقاتی برای اعمال هر دوی این روش ها انجام شده است که از دقت و سرعت مناسب نیز برخوردارند [۱۱۹-۱۲۱]. همچنین، روش هایی برای جایگزینی این دو روش موجود پیشنهاد شده است. به طور مثال، تابع تراکم آنتروپی متقاطعی[47] که در مرجع [۱۲۲] برای جایگزینی CTC و روش های توجه، طراحی شده، علاوه بر آن که عملکرد رقابتی خوبی در مقابل این دو روش داشته است، پیاده سازی سریع تری دارد و حجم حافظه کمتری مورد نیاز است. در ادامه، هر دو روش CTC و مبتنی بر ساز و کار توجه توضیح داده می شوند.

۲-۴-۱ روش های مبتنی بر CTC

در روش CTC، تصویر ورودی به تعدادی گام زمانی تقسیم می شود و تعداد ویژگی های ثابتی برای هر یک از این گام های زمانی به دست می آید. سپس، نویسه مربوط به هر گام زمانی، پیش بینی می شود. از این رو، در این روش، پیش بینی دنباله های با طول های مختلف انجام می شود؛ با این تفاوت که هر تصویر با توجه به طول متن موجود در آن، به تعداد گام های زمانی متفاوتی تقسیم می شود. نکته کلیدی در روش CTC، بازشناسی یک نویسه در هر گام زمانی از تصویر متن ورودی است. در این مرحله، دنباله ای از نویسه ها به همراه یک شبه نویسه با نام فاصله تولید می شود. از شبه نویسه فاصله برای حالت هایی استفاده می شود که دو نویسه تکراری در یک کلمه پشت سر هم می آیند، اما وجود هر دو نویسه برای آن کلمه نیاز است. به طور مثال، در کلمه pizza، که دارای دو نویسه z پشت سر هم است. در این حالت ها، بین دو نویسه تکراری، وجود یک شبه نویسه فاصله ضروری است. در نهایت، برای بدست آوردن متن نهایی، ابتدا حروف تکراری پشت سر هم و سپس فاصله ها حذف می شوند. شکل ۵، ساختار پایه یک سامانه بازشناسی حروف مبتنی بر CTC را نشان می دهد.



شکل ۵. مدل پایه CNN، LSTM و CTC برای بازشناسی حروف [۱۱۳]

[42] Multi-dimensional recurrent neural networks (MDRNNs)
[43] Bidirectional LSTM
[44] Sequence-to-sequence
[45] Connectionist Temporal Classification
[46] Attention mechanism
[47] Aggregation cross-entropy function

CTC در مرجع [۹۱] برای آموزش شبکه های بازگشتی در برچسب زنی مستقیم دنباله های جداسازی نشده، معرفی شد. CTC موفقیت های خوبی در زمینه های بازشناسی صوت [۱۲۳, ۱۲۴] و بازشناسی متون دستنوشته برخط [۱۰۸] داشته است. به دنبال این موفقیت ها، مراجع [۹۲, ۱۱۳, ۱۲۵]، از اولین مراجعی بودند که از CTC برای بازشناسی متون در منظره استفاده کردند و پس از آن ها، کارهای بسیاری [۱۲۶, ۱۲۷] در بازشناسی حروف مبتنی بر CTC انجام شد. در مسأله بازشناسی حروف، CTC، ویژگی های تولید شده به وسیله شبکه های کانولوشنی و بازگشتی را به دنباله ای از نویسه ها تبدیل می کند. این لایه، از برچسب با بیشترین احتمال برای بازشناسی برچسب هر گام زمانی ورودی استفاده می کند. سپس، برچسب های تمام گام های زمانی را برای تولید متن نهایی ادغام می کند.

یکی از زمینه های بالقوه ای که برای تحقیقات بیشتر وجود دارد، اصلاح ساختار CTC برای کار با متن های دارای اعوجاج های فضایی نظیر نوشته های موجود در منظره است. با توجه به بهبود نسبی که در مرجع [۱۲۸] برای مدل دو بعدی CTC نسبت به CTC استاندارد ارائه شده است، تحقیقات بیشتر در این زمینه می توانند ارزشمند باشند.

۱-۴-۲ روش های مبتنی بر توجه

مدل توجه، همانطور که از نام آن مشخص است، سعی می کند توجه شبکه عصبی را به قسمت خاصی از دنباله جلب کند که بیشترین اطلاعات را برای تخمین خروجی داشته باشد و به قسمت های کم اهمیت دنباله ورودی، توجه کم‌تری نشان می دهد. ساز و کار توجه در یادگیری عمیق، می تواند به طور کلی به عنوان یک بردار وزن تفسیر شود که برای پیش بینی یک المان نظیر یک پیکسل در یک تصویر یا یک کلمه در یک جمله استفاده می شود. در حقیقت، توجه، یک لایه پس خور[۴۸] با وزن های قابل آموزش است که میزان وابستگی بین المان های مختلف یک دنباله را نشان می دهد.

ساز و کار توجه در ابتدا در زمینه پردازش زبان طبیعی[۴۹] (NLP) پیشنهاد شد و هنوز هم در این حوزه مورد توجه محققان قرار دارد. در ادامه، تحقیقات کلیدی از بدو مطرح شدن ساز و کار توجه تا کنون ذکر می شوند. اگر چه هر یک از این مراجع ساز و کار توجه را در کاربرد خاصی استفاده کرده‌اند، اما مراجع مختلفی برای استفاده از مدل های پیشنهادی در این مقاله ها برای کاربردهای دیگر، منتشر شده است.

مدل های توجه، در راستای بهبود مدل‌هایی به نام دنباله به دنباله[۵۰] یا شبکه عصبی بازگشتی رمز گذار-رمز گشا[۵۱] [۱۲۹, ۱۳۰] ارائه شدند که برای نگاشت یک دنباله در ورودی به یک دنباله در خروجی استفاده می شوند. مدل های توجه در ابتدا، برای حل مسأله تراز[۵۲] و ترجمه ماشینی در مراجع [۱۳۱, ۱۳۲] پیشنهاد شدند. تراز بندی در مسأله ترجمه ماشینی، مشخص می کند که کدام قسمت از دنباله ورودی مربوط به هر کلمه در خروجی است، در حالیکه ترجمه، فرآیند استفاده از اطلاعات مرتبط برای انتخاب خروجی مناسب است.

مدل بعدی، توجه بصری است که برای اولین بار در مرجع [۱۳۳] برای بسط روش های دنباله به دنباله ارائه شد. این چارچوب سعی می کند تصویر ورودی و کلمه خروجی را در مسأله برچسب گذاری تصویر[۵۳]، تراز کند. این چارچوب، یکی از اولین تلاش ها برای اعمال ساز و کار توجه در کاربرد دیگری غیر از ترجمه ماشینی است. تصویر به یک شبکه کانولوشنی داده می شود تا ویژگی های مناسب استخراج شوند. سپس، این ویژگی ها به یک شبکه کد گشا (شبکه بازگشتی با ساز و کار توجه) داده می شوند تا در هر مرحله، یک کلمه از توضیح عکس را تولید کند. در حقیقت، کد گشا در هر مرحله برای تولید کلمه، از ساز و کار توجه استفاده می کند تا روی قسمت مناسبی از تصویر توجه کند.

توجه سلسله مراتبی[۵۴]، ساختار دیگری از ساز و کار توجه است که در مرجع [۱۳۴] پیشنهاد شد. آن ها نشان دادند که توجه می تواند در سطوح مختلفی مورد استفاده قرار گیرد. این ساختار برای مسأله طبقه بندی اسناد استفاده شده است. این شبکه، تفسیر سلسله مراتبی از نتایج را ممکن می سازد. بدین صورت که مشخص می کند چه جمله ای در طبقه بندی یک سند مهم است و چه قسمتی از این جمله، یعنی کدام کلمات، در آن جمله برجسته تر هستند.

معماری شبکه عصبی مبدل[۵۵] [۱۳۵] یکی از بزرگ ترین پیشرفت های دهه در زمینه NLP است. مشابه شبکه های عصبی بازگشتی، مبدل ها برای مدیریت داده های دنباله دار

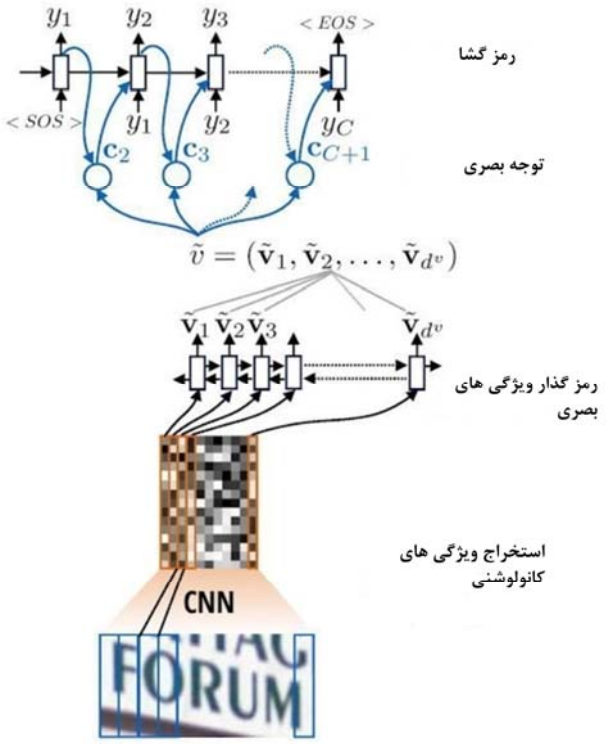
[48] Feed-forward
[49] Natural language processing

[50] Sequence to sequence (seq2seq)
[51] RNN Encoder-Decoder Model
[52] Alignment
[53] Image captioning
[54] Hierarchical attention
[55] Transformer neural network

نظیر ترجمه و خلاصه سازی متن طراحی شده اند. با این وجود، بر خلاف شبکه های بازگشتی، مبدل ها نیازی ندارند که این داده ها را به ترتیب حضورشان در دنباله پردازش کنند. به واسطه این ویژگی، امکان موازی سازی را بیشتر از شبکه های بازگشتی فراهم می آورد و در نتیجه، زمان آموزش کمتر می شود. معماری مبدل، عملیات کانولوشنی و بازگشتی را کنار می گذارد و ساز و کار توجه چندگانه[۵۶] را جایگزین آن ها می کند. توجه چندگانه، در اصل چندین لایه توجه است که به طور مشترک، بازنمایی های مختلف را از مکان های مختلف یاد می گیرند.

مدل های الهام گرفته شده از مدل مبدل، نظیر BERT [۱۳۶] رکود شکنی را در بسیاری از کارهای NLP نشان داد. با توجه به آنچه که گفته شد، نویسندگان مرجع [۱۳۷] ادعا کردند که می توان از مبدل برای کارهای بینایی ماشین استفاده کرد. این ادعا شاید کمی قدیمی به نظر برسد، زیرا پیش از این، ساز و کار توجه برای کارهای مربوط به تصویر در مرجع [۱۳۳] پیشنهاد شده بود. با این وجود، ادعای این مرجع [۱۳۷] یک انقلاب محسوب می شود؛ زیرا مبدل به جای تکمیل لایه های کانولوشنی، عملا جایگزین آن ها می شود. همچنین، این مبدل بصری زمانی که با داده های کافی آموزش ببیند، از مدل های پیشرفته CNN در مقیاس بزرگ نیز بهتر عمل می کند. این مطلب می تواند به این معنی باشد که دوران طلایی CNN ها که سال ها به طول انجامید، مشابه RNNها به پایان خود نزدیک می شود.

همانطور که گفته شد، اگرچه ساز و کار توجه در مقاله های مرجع برای کاربردهای خاصی پیشنهاد شدند، این روش ها در مراجع مختلفی برای سایر کاربردها نیز بسط داده شده اند. در این راستا، تحقیقات مختلفی برای بازشناسی حروف انجام شده است که از جمله آن ها می توان به مراجع [۱۳۸-۱۴۵] در بازشناسی دستنوشته و مراجع [۱۰۲, ۱۰۵-۱۰۷, ۱۴۶-۱۵۱] برای بازشناسی حروف در منظره اشاره کرد. در ادامه، تعدادی از این مراجع مرور می شوند. شکل ۶، یک ساختار مبتنی بر توجه را برای بازشناسی حروف ارائه می دهد. این ساختار را می توان یک مدل CRNN در نظر گرفت که در دنباله آن به جای استفاده از CTC از ساز و کار توجه استفاده شده است.



شکل ۶. یک ساختار بازشناسی نوری حروف مبتنی بر توجه[۵۷]

همانطور که گفته شد، باهادانا [۱۳۱] اولین فردی بود که ساختار رمز گشای مبتنی بر توجه را برای کاربرد ترجمه ماشینی پیشنهاد داد. هدف از این کار، معطوف کردن توجه شبکه به ورودی های مرتبط و موثر برای یافتن خروجی مناسب در هر گام زمانی است. از سوی دیگر در مرجع [۱۵۲]، روشی به نام شبکه مبدل فضایی[۵۸] معرفی شده است که می تواند در ساختار یک شبکه عصبی استاندارد نظیر CNN یا یک شبکه تمام متصل[۵۹] استفاده شود. این ماژول، یک تصویر ورودی را از نظر ساختار فضایی اصلاح می کند تا شبکه نسبت به تغییرات فضایی نظیر چرخش، کجی و سایز تصویر، پایدار بماند. نویسندگان مرجع [۱۰۴]، با استفاده از این ماژول و با الهام از روش باهادانا، الگوریتمی برای بازشناسی تصاویر متن دار موجود در منظره و تصاویر کلمه هایی که دچار اعوجاج شده اند، پیشنهاد داده است. در مرجع [۱۵۱]، ساختاری شبیه روش CRNN پیشنهاد شده است که در آن به جای استفاده از CTC، از ساز و کار توجه استفاده شده که از مدل توجه دنباله به دنباله [۱۳۱] الهام گرفته شده است. مدل پیشنهادی روی مجموعه داده FSNS که مجموعه داده ای لاتین از نوشته ها و تابلوهای موجود در خیابان است، دقت بالایی را کسب کرده است.

[56] Multi-head attention

[57] https://nanonets.com/blog/attention-ocr-for-text-recogntion/

[58] Spatial transformer network

[59] Fully connected

مدل توجه بازگشتی[60] (RAM) [۱۴۴] روش دیگری است که با استفاده از یادگیری تقویتی[61] سعی می کند عملکرد چشم انسان را تقلید کند. زمانی که انسان با یک صحنه جدید مواجه می شود، با یک نگاه اجمالی، نواحی خاصی از تصویر توجه او را به سمت خود جلب می کنند و یک سری اطلاعات را در همان نگاه اجمالی کسب می کند. با گذر زمان، توجه چشم به جزئیات مهم تصویر معطوف می شود. در مدل RAM، یک بردار به نام glimpse (نگاه اجمالی) تعریف می شود که ویژگی های تصویر را در اطراف یک مکان خاص استخراج می کند. تصویر در اندازه های مختلف در اطراف یک مرکز مشترک بریده می شود و بردارهای glimpse با ویژگی های برجسته و موثر از هر نسخه برش خورده ایجاد می شوند. این بردارهای glimpse، صاف[62] می شوند و از یک شبکه glimpse مبتنی بر توجه بصری عبور می کنند. سپس، بردارهای glimpse به یک شبکه مکان[63] منتقل می شوند که از یک RNN برای پیش بینی قسمت بعدی تصویر که باید مورد توجه قرار بگیرد، استفاده می کند. این مکان، ورودی بعدی برای شبکه glimpse است. به تدریج، این مدل، قسمت های دیگری از تصویر را کاوش می کند و در هر مرتبه، پس انتشار[64] را اعمال می کند تا ببیند آیا اطلاعات قبل بردارهای glimpse برای دستیابی به دقت مناسب، کافی هستند یا خیر.

اگرچه مدل RAM توانایی طبقه بندی تصاویر مربوط به ارقام یا حروف گسسته را به خوبی دارد و می تواند اعداد و حروف را در تصاویر شلوغ[65] شناسایی کند، این مدل برای طبقه بندی چندین هدف در یک تصویر، مناسب نیست. از این رو، مدل مبتنی بر توجه دیگری به نام DRAM[66] در مرجع [۱۴۵] پیشنهاد شده است. DRAM دارای ۵ قسمت است: شبکه glimpse، شبکه بازگشتی، شبکه زمینه[67]، شبکه طبقه بندی و شبکه انتشار[68]. بزرگ‌ترین تفاوت بین RAM و DRAM این است که DRAM از دو شبکه بازگشتی پشت سر هم استفاده می کند. اولین واحد، وظیفه طبقه بندی را بر عهده دارد و دومین واحد، مسئول تعیین مکان glimpse بعدی است. حالت اولیه واحد بازگشتی دوم به وسیله شبکه زمینه تولید می شود (این در حالی است که در روش RAM، مکان اولیه به صورت تصادفی انتخاب می شد). شبکه زمینه، از سه لایه کانولوشنی تشکیل شده است که وظیفه نمونه کاهی[69] تصویر ورودی را دارد. این تصویر نمونه کاهی شده، برخی نکات را برای دادن مختصات مکان ارائه می دهد. روش RAM تلاش می کند تا یک برچسب را برای یک تصویر تولید کند، در حالیکه DRAM یک دنباله از برچسب ها را برای چند هدف در تصویر تولید می کند.

مرجع [۱۴۱]، روی بازشناسی متون دستنویس لاتین با استفاده از معماری دنباله به دنباله مبتنی بر توجه کار کرده است. این مرجع، ساز و کار مبتنی بر توجه را به شش گروه اصلی به نام های توجه مبتنی بر محتوا[70]، توجه مبتنی بر جریمه[71]، توجه مبتنی بر مکان[72]، توجه یکنواخت[73]، توجه تکه به تکه[74] و توجه ترکیبی[75] تقسیم کرده و عملکرد هر یک از آن ها را روی چند مجموعه داده از متون لاتین آزمایش و مقایسه کرده است.

مرجع [۱۳۸]، شبکه های توجه مبتنی بر نویسه (به جای مبتنی بر کلمه) را برای بازشناسی متون دستنویس لاتین استفاده کرده است. به عبارت دیگر، شبکه های رمز گذار رمز گشا، روی دنباله ای از نویسه ها به جای کلمات، آموزش می بینند. توسعه مدل های آگاه از نویسه[76] (یعنی مدل هایی که خطوط ورودی و خروجی را به عنوان دنباله هایی از نویسه ها به جای کلمات در نظر می گیرند) می تواند بسیار موثر باشد؛ زیرا این مدل ها می توانند درباره کلمه هایی که تا کنون ندیده اند استنتاج کنند. به علاوه، مدل های آگاه از نویسه، به دامنه زیادی از واژگان نیاز ندارند، زیرا تنها نویسه ها به طور واضح مدل سازی می شوند.

۳. مروری بر تحقیقات انجام شده با یادگیری عمیق برای بازشناسی نوری نویسه ها در خط فارسی، عربی و اردو

اگرچه در بخش های مختلف مقاله، به پژوهش های انجام شده در حوزه بازشناسی نوری نویسه ها با استفاده از الگوریتم های یادگیری عمیق در رسم الخط های با نوشتار پیوسته، به خصوص فارسی، عربی و اردو اشاراتی شده است، در این

[60] Recurrent Attention Model
[61] Reinforcement learning
[62] Flattened
[63] Location network
[64] Back propagation
[65] Cluttered
[66] Deep recurrent attention model
[67] Context network
[68] Emission network
[69] Down-sampling
[70] Content-based attention
[71] Penalized attention
[72] Location-based attention
[73] Monotonic attention
[74] Chunkwise attention
[75] Hybrid attention
[76] Character-aware models

بخش قصد داریم به طور ویژه به مرور این آثار بپردازیم، تا کار برای علاقه‌مندان به پژوهش در این حوزه، آسان تر باشد.

گروهی از تحقیقات، از شبکه های عمیق کانولوشنی برای بازشناسی نویسه هایی که به صورت گسسته نگارش شده اند، استفاده کرده اند. شبکه های کانولوشنی عموما برای طبقه بندی کلاس های از پیش تعریف شده استفاده می شوند. بنابراین، کاربرد این شبکه ها در بازشناسی نوری نویسه ها، به شناسایی ارقام، حروف گسسته و نهایتا زیر کلمه هایی که شکل مشابهی در زبان های با خط پیوسته دارند، محدود می شود. به طور مثال، مرجع [۱۵۳] از وزن های آموزش دیده مدل های AlexNet و GoogleNet که روی مجموعه تصاویر ImageNet آموزش دیده اند، برای طبقه بندی ۵۴ حرف و عدد دستنویس مربوط به خط اردو استفاده کرده است. در مرجع [۱۵۴]، ابتدا با استفاده از یک ساختار کانولوشنی، ویژگی های نویسه های گسسته عربی استخراج می شوند. سپس، از طبقه بند ماشین بردار پشتیبان، در لایه آخر برای طبقه بندی این نویسه ها استفاده شده است. روش تقریبا مشابهی برای شناسایی نویسه های دستنویس فارسی در مرجع [۱۵۵] ارائه شده است. شناسایی لیگاتورهای[۷۷] اردو با استفاده از شبکه عصبی کانولوشنی در مرجع [۱۵۶] دنبال شده است. نویسندگان در مرجع [۱۵۷]، به بررسی دقت طبقه بندی اعداد رنگی دستنویس فارسی پرداخته اند. در این مقاله نیز با استفاده از شبکه کانولوشنی و یک مجموعه داده شامل ۱۳،۳۳۰ تصویر رنگی از ارقام ۰-۹ استفاده شده است.

مرجع [۱۵۸]، از ساختار LSTM دو سویه چند جهتی[۷۸] (MD-BLSTM) و لایه CTC استفاده کرده و بازشناسی متون چاپی عربی را بدون نیاز به جداسازی انجام داده است. همچنین، مرجع [۱۵۹]، از ساختار CNN-BLSTM- CTC برای بازشناسی متن دستنویس عربی استفاده کرده است. مرجع [۱۶۰]، نیز از یک ساختار مشابه برای بازشناسی متن های عربی موجود در منظره استفاده کرده است. این مرجع، برای بررسی کیفیت روش پیشنهادی، کار خود را روی مجموعه داده های ALIF و ACTIV و همچنین یک مجموعه داده مصنوعی دارای ۲۰۰۰ کلمه آزمایش کرده است. مقایسه نتیجه حاصل از این روش با روش tesseract، نشان دهنده کارکرد مناسب این روش است.

استفاده از روش های مبتنی بر توجه در بازشناسی متون فارسی، عربی و اردو محدود است و از این رو فضای تحقیق در این حوزه بسیار باز است. نکته ای که در اینجا قابل ملاحظه است، این است که با توجه به پیوستگی حروف در این سه خط، می توان از روش های مبتنی بر توجهی که برای بازشناسی متون دستنویس لاتین معرفی شده اند، استفاده کرد؛ زیرا در نوشته های دستنویس لاتین، پیوستگی حروف تا حد زیادی وجود دارد. مرجع [۱۶۱] یک معماری رمزگذار/رمزگشای مبتنی بر توجه را برای بازشناسی جمله های اردو معرفی کرده است. در مرحله رمزگذاری، ویژگی های سطح بالا از تصویر ورودی به وسیله یک DenseNet استخراج می شوند. در مرحله رمزگشایی، این ویژگی های سطح بالا، نویسه به نویسه رمزگشایی می شوند تا جمله خروجی به دست آید. زمان رمزگشایی یک نویسه در یک بازه زمانی، به جای کل تصویر، تنها به قسمت های مرتبط تصویر تمرکز می شود. به علاوه، فرایند آموزش برای رمزگذار و رمزگشا به صورت همزمان انجام می پذیرد.

جدول ۱، خلاصه ای از تحقیقات انجام شده برای بازشناسی متن های فارسی، عربی و اردو را که از شبکه های ژرف استفاده کرده اند، نشان می دهد.

[77] Ligatures
[78] Bi-direction long short-term memory

جدول ۱. خلاصه ای از روش های بازشناسی مبتنی بر الگوریتم های ژرف در خط فارسی، عربی و اردو

مرجع	سال	مدل	مجموعه داده	نوع	اندازه	دقت
[۱۶۲]	۲۰۱۶	TDNN	مجموعه داده اختصاصی	حروف کلمات	۰۹۰،۶ ۰۸۰،۱	۹۸%/۵ ۹۶%/۹۰
[۱۵۴]	۲۰۱۶	CNN-SVM	HACDB IFN/ENIT	حروف	۶،۶۰۰ ۲۱۲،۲۱۱	۹۴%/۱۳ ۹۲%/۹۵
[۱۶۳]	۲۰۱۷	MLP CNN	CMATERDB[۷۹]	ارقام	۳،۰۰۰	۹۳%/۸ ۹۷%/۴
[۱۶۴]	۲۰۱۷	RBM-CNN	CMATERDB	ارقام	۳،۰۰۰	۹۸%/۵۹
[۱۶۵]	۲۰۱۷	CNN	مجموعه داده اختصاصی	حروف	۱۶،۸۰۰	۹۴%/۹
[۱۶۶]	۲۰۱۷	CNN	مجموعه داده اختصاصی	زیر کلمه	۹۶،۲۸۹	۸۳%/۵
[۱۶۷]	۲۰۱۷	CNN (VGG)	ADBase HACDB	ارقام حروف	۷۰،۰۰۰ ۶،۶۰۰	۹۹%/۵۷ ۹۷%/۳۲
[۱۵۵]	۲۰۱۷	CNN(LeNet-5)	HODA	حروف+ارقام	۸۸،۳۵۱+۸۰،۰۰۰	۹۷%/۱
[۱۵۶]	۲۰۱۷	CNN	مجموعه داده اختصاصی	لیگاتور	۵۵،۰۰۰	۹۵%
[۱۶۸]	۲۰۱۷	CNN	AHCD AIA9K	حروف حروف	۱۶،۸۰۰ ۸،۷۳۷	۹۷%/۶ ۹۴%/۸
[۱۵۹]	۲۰۱۸	CNN-BLSTM-CTC	IFN/ENIT	کلمات	۲۶،۴۵۹	۹۲%/۲۱

[79] https://code.google.com/archive/p/cmaterdb/downloads

[۱۶۹]	۲۰۱۸	CNN	MADBase MNIST HODA Urdu DHCD	ارقام	۷۰،۰۰۰ ۷۰،۰۰۰ ۰۰۰،۸۰ ۸،۵۰۰ ۲۰،۰۰۰	۹۹%/۳۰ ۹۹%/۴۳ ۹۹%/۸۲ ۹۹%/۳۲ ۹۹%/۷۳
[۱۷۰]	۲۰۱۸	ساز و کار توجه	KHATT	خط	۹،۳۲۷	۹۳%/۰۷
[۱۷۱]	۲۰۱۹	CNN	CMATERDB	ارقام	۳،۰۰۰	۹۹%/۴
[۱۷۲]	۲۰۲۰	CNN	Hijja AHCD	حروف	۴۷،۴۳۴ ۱۶،۸۰۰	۹۷% ۸۸%
[۱۵۳]	۲۰۲۰	CNN (AlexNet) CNN(GoogleNet)	IFHCDB	ارقام+حروف	۱۷،۷۴۰+۵۲،۳۸۰	۹۶%/۳ ۹۴%/۷
[۱۶۱]	۲۰۲۰	ساز و کار توجه	مجموعه داده اختصاصی	حرف کلمه	۷۸،۸۷۰ ۲۸۳،۶۶۴	۷۷%/۰۵ ۴۳%/۳۵

۴. پایگاه های داده

پیدایش مجموعه های داده متنوع، موجب ایجاد چالش های جدید در مبحث بازشناسی نوری حروف می شود. از سوی دیگر، روش های یکپارچه ارزیابی دقت، موجب رعایت انصاف برای مقایسه روش های مختلف می شود. در این بخش، مجموعه های داده پرکاربرد و همچنین، معیارهای ارزیابی کارایی روش های بازشناسی حروف، مطرح می شوند.

در حال حاضر، مجموعه های داده مناسبی برای هر سه کاربرد، متون دستنویس، چاپی و داده های متنی در منظره وجود دارد. بر مبنای نحوه جمع آوری داده ها، مجموعه های داده موجود برای باشناسی حروف، به دو دسته مجموعه های داده واقعی و مصنوعی تقسیم می شوند. مجموعه های داده واقعی، از پویش اسناد، دستنوشته ها و تصاویر گرفته شده از نوشته های موجود در منظره به وجود می آیند. این در حالی است که برای ساخت مجموعه های داده مصنوعی از متن های آماده استفاده می شود و با اعمال انواع قلم ها و بعضا پس زمینه های متنوع به صورت تصادفی، تصویری از هر یک از خطوط یا کلمات متن تهیه می شود. متن صحیح متناظر با این تصویر نیز که از قبل موجود است. به عبارت دیگر، روندی بر عکس روند معمول، برای ایجاد مجموعه های داده مصنوعی به کار می رود. در روش معمول، متن موجود در تصاویر اسکن شده به صورت دستی یا با استفاده از موتورهای بازشناسی متن، بازشناسی می شد که البته نیاز به اصلاح فراوان داشت. در اینجا، با استفاده از متن صحیح، اقدام به تولید تصویر می شود که کاری به مراتب ساده تر است.

توسعه مجموعه های داده استاندارد، مورد توجه بسیاری از محققین از ابتدای سال ۱۹۹۰ قرار گرفت [۱۷۳]. مهم ترین و پرکاربردترین مجموعه های داده واقعی دستنویس عبارتند از

[۱۷۴-۱۷۶]IAM، [۱۷۷] RIMES، [۱۷۸] NIST، [۱۷۹] MNIST، [۱۸۰-۱۸۵] CENPARMI، [۱۸۶] CEDAR ، [۱۸۶] UNIPEN، [۱۸۶] ETL9 و [۱۸۷] PE92 [۱۸۸]. اگرچه اکثر این مجموعه داده ها با استفاده از زبان های مبتنی بر الفبای لاتین توسعه یافته اند، توسعه مجموعه داده ها در زبان های چینی [۱۸۹]، کره ای [۱۸۸]، عربی [۴۵, ۱۸۰, ۱۹۰-۱۹۳]، فارسی [۱۸۲, ۱۸۵, ۱۹۴-۱۹۶] و هندی [۱۹۷] نیز مد نظر قرار گرفته است. همچنین، در سال های اخیر، ایجاد مجموعه های داده چند زبانه [۱۹۸, ۱۹۹] مورد توجه قرار گرفته است. این مجموعه های داده دستنویس، انواع متنوعی از نمونه ها را شامل اعداد دستنویس [۱۷۸-۱۸۰, ۱۹۲, ۲۰۰]، حروف [۱۸۶, ۱۸۸, ۲۰۱-۲۰۳]، کلمات [۱۸۰, ۱۸۶, ۱۹۰, ۱۹۲, ۱۹۴, ۲۰۰] و جمله کامل [۱۷۵, ۱۷۷, ۱۸۹, ۱۹۸, ۱۹۹]، در بر می گیرند. از جمله کارهای انجام شده برای تهیه مجموعه های داده دستنویس، می توان به مراجع [۲۰۴-۲۰۶] اشاره کرد.

مجموعه های داده چاپی به طور معمول، از پویش صفحات روزنامه ها و کتاب ها به دست می آیند. به طور مثال، مرجع [۱۱۴] از تصاویر پویش شده روزنامه های قدیمی و امروزی و همچنین فاکتورهای به زبان آلمانی استفاده کرده است. این اسناد پس از تصحیح کجی[80] (در صورت وجود) با استفاده از موتور بازشناسی نوری حروف Tesseract [۲۰۷] بازشناسی شده و پس از آن به صورت دستی، خطاهای ناشی از Tesseract اصلاح شده اند. همچنین، مجموعه های داده APTID/MF [۲۰۸]، PATDB [۲۰۹] و ATID [۲۱۰] از مجموعه های داده چاپی عربی هستند. مجموعه داده ALTID [۲۱۱] یک مجموعه داده چاپی دوزبانه از زبان های عربی و لاتین است. در گردآوری مجموعه های داده چاپی، از روش های مصنوعی نیز استفاده می شود. به طور مثال، APTI [۲۱۲] یک مجموعه داده چاپی مصنوعی با ۴۵،۳۱۳،۶۰۰ کلمه و بیش از ۲۵۰ میلیون نویسه عربی است که از نوشتن ۱۱۳،۲۴۸ کلمه عربی با قلم ها و اندازه قلم های مختلف و همچنین با سبک های مختلف قلم از جمله سبک های با حروف درشت/کج[81] نوشته شده است. مجموعه داده شتر[82] نیز یک مجموعه داده شامل ۱۲۰،۰۰۰ کلمه فارسی است. کلمات این مجموعه داده از سایت های ویکی پدیای فارسی و گنجور استخراج شده اند.

مجموعه های داده موجود برای متون در منظره را نیز می توان بر اساس واقعی یا مصنوعی بودن آن ها، تقسیم بندی کرد. از جمله مهم ترین مجموعه های داده مصنوعی در این زمینه، می توان به Synth90k [۹۸]، SynthText [۲۱۳]، VerisimilarSynthesis [۲۱۴]، UnrealText [۲۱۵] اشاره کرد که همه آن ها به زبان انگلیسی هستند. مجموعه های داده ALIF [۲۱۶] و ACTIV [۲۱۷] را می توان به عنوان دو مجموعه داده مصنوعی عربی نام برد که شامل فریم هایی از ویدئوهای مربوط به کانال های خبری عربی زبان هستند. مجموعه داده ALIF شامل ۶۵۳۲ تصویر خط متن جداسازی شده از ۵ کانال خبری معروف عربی است. مجموعه داده ACTIV، بزرگ تر از مجموعه داده ALIF است و شامل ۲۱۵۲۰ تصویر از کلمات از کانال های خبری عربی می باشد. مجموعه های داده واقعی موجود در منظره، خود به سه دسته کلی قاعده مند، بی قاعده و چند زبانه تقسیم می شوند. در مجموعه های داده قاعده مند، اکثر نمونه های متنی به صورت افقی هستند و ممکن است تنها قسمت کوچکی از آن ها مخدوش باشد. مجموعه های داده IIIT5K-Words (IIIT5K) [۲۱۸]، Street View Text (SVT) [۲۱۹]، ICDAR 2003 (IC03) [۲۲۰]، ICDAR 2011 (IC11) [۲۲۱]، ICDAR 2013 (IC13) [۲۲۲] و Street View House Number (SVHN) [۲۲۳] از مجموعه های داده قاعده مند با حروف لاتین هستند. تعداد مجموعه دادهای حروف در منظره برای زبان های فارسی و عربی بسیار کم است. از جمله مجموعه های داده عربی تصاویر متن دار در منظره می توان به ARASTI [۲۲۴] اشاره کرد که شامل ۱۶۸۷ تصویر، ۱۲۸۰ کلمه های عربی جداسازی شده و ۲۰۹۳ حرف عربی جداسازی شده در منظره است. ARASTEC [۲۲۵]، یکی دیگر از این مجموعه داده ها با ۲۶۰ تصویر متن دار در منظره است. حروف موجود در تصاویر این مجموعه داده به صورت دستی جداسازی شده اند که منجر به ایجاد ۱۰۰ کلاس از حروف در منظره با توجه به جایگاه های مختلف حرف در کلمه شده است. در مجموعه های داده بی قاعده، اکثر نمونه های متنی با وضوح پایین و به صورت منحنی یا دارای پرسپکتیو هستند. قلم های مختلف و الگوهای مخدوش متن های بی قاعده، چالش های اضافی را برای این دسته از تحقیقات به وجود می آورد. از جمله مجموعه های داده لاتین بی قاعده، می توان به مجموعه های داده StreetViewText-Perspective (SVT-P) [۲۲۶]، CUTE80 (CUTE) [۲۲۷]، ICDAR 2015 (IC15) [۲۲۸]، COCO-Text [۲۲۹] و Total-Text [۲۳۰] اشاره کرد. متن های دو زبانه و چند زبانه در شهرهای توسعه یافته به وفور پیدا می شوند. در مرجع [۲۳۱]، تصاویر منظره

[80] deskewing
[81] Bold/italic
[82] Camel

حاوی زبان های عربی و انگلیسی گرد آوری شده است. مراجع [۲۳۲-۲۳۴]، از جمله مجموعه های داده گردآوری شده دو زبانه لاتین و چینی است. همچنین، مرجع [۲۳۵] یک مجموعه داده ده زبانه از زبان های عربی، بنگلادشی، چینی، دواناگری[۸۳]، انگلیسی، فرانسوی، آلمانی، ایتالیایی، ژاپنی و کره ای است.

با توجه به اینکه روش های مبتنی بر الگوریتم های ژرف بازشناسی حروف، مستقل از زبان هستند و یک ساختار شبکه ای یکسان را می توان روی زبان های مختلف، آموزش داد و از آن نتیجه گرفت، آموزش شبکه با استفاده از مجموعه های داده چند زبانه می تواند بسیار موثر باشد. در حال حاضر، تعداد مجموعه های داده چند زبانی محدود است [۱۹۸, ۱۹۹] و تا آنجا که می دانیم، مجموعه داده چند زبانی فارسی-انگلیسی غنی اعم از متنون چاپی یا متن های موجود در منظره، وجود ندارد.

از آنجا که تاکید این مقاله به رسم الخط فارسی، عربی و اردو است، اطلاعات کامل راجع به مجموعه داده های موجود در این سه زبان در جدول ۲ درج شده است.

[83] Devanagari

جدول ۲. ویژگی های مجموعه داده های پر کاربرد برای بازشناسی نویسه ها در رسم الخط فارسی، عربی و اردو.

مرجع	نام مجموعه داده	زبان	چاپی/دستنویس/ منظره	رقم/حرف/ کلمه/جمله	تعداد نمونه ها
[۱۹۵]	Hoda	فارسی	دستنویس	رقم/ حرف	۱۰۲،۳۵۲/ ۸۸،۳۵۱
[۲۳۶]	شتر	فارسی	چاپی (مصنوعی)	کلمه	۱۲۰،۰۰۰
[۲۳۷]	FHT	فارسی	دستنویس	جمله/ کلمه/ زیر کلمه	۸،۰۵۰/ ۱۰۶،۶۰۰/ ۲۳۰،۱۷۵
[۲۳۸]	IFHCDB	فارسی	دستنویس	حرف/ رقم	۵۲،۳۸۰ ۱۷،۷۴۰/
[۱۸۵]	CENPARMI	فارسی	دستنویس	تاریخ/ رقم (فارسی)/ رشته ای از ارقام/ حرف/ کلمه (در چک های بانکی)/ رقم (انگلیسی، که توسط افراد فارسی زبان نوشته شده اند)	۱۷۵/ ۱۸،۰۰۰/ ۷،۳۵۰/ ۱۱،۹۰۰/ ۸،۵۷۵/ ۳،۵۰۰
[۱۲]	Sadri	فارسی	دستنویس	تاریخ (به رقم)/ تاریخ (به حروف)/ رقم (فارسی)/ رشته ای از ارقام/ حرف/ سمبل ها و علائم نگارشی/ علائم ریاضی/ کلمه های پر کاربرد فارسی در زمینه اسامی اشخاص،	۲،۵۰۰/ ۲،۰۰۰/ ۹۷،۱۲۴/ ۴۵۰/ ۴۳،۰۰۰/ ۱۱،۵۰۰/ ۱۶،۰۰۰/ ۷۰،۰۰۰/

					/۵۰۰ محصولات تجاری، اعداد اصلی و ترتیبی/کلمات پرکاربرد چك های بانکی، اسمی ماه های شمسی و قمری، اسامی چند شهر ایران/ متن (شامل یك پاراگراف)
[۱۹۰]	IFN/ENIT	عربی	دستنویس	کلمه/ زیر کلمه/ حرف لیگاتور	/۲۶،۴۵۹ /۱۱۵،۵۸۵ /۲۱۲،۲۱۱ ۴،۱۴۳
[۱۸۰]	Arabic Cheque	عربی	دستنویس	رقم/ زیر کلمه	/۱۵،۱۷۵ ۲۹،۴۹۸
[۲۳۹]	ADBase/ MADBase Datasets[84]	عربی/ عربی- انگلیسی	دستویس	رقم	۷۰،۰۰۰
[۲۴۰]	HACDB[85]	عربی	دستنویس	حرف	۶،۶۰۰
[۲۴۱]	KHATT	عربی	دستنویس	حرف/ کلمه/ خط/ تك گرام/ بایگرام/ ترایگرام	/۵۸۹،۹۲۴ ۱۶۵،۸۹۰ /۹،۳۲۷ /۲۸،۷۴۳ /۸۷،۸۵۰ ۱۰۱،۰۸۱
[۲۱۶]	ALIF	عربی	منظره (مصنوعی)	حرف/ زیر کلمه/ کلمه/	/۸۹،۸۱۹ /۵۲،۴۱۰ /۱۸،۰۴۱

[84] http://datacenter.aucegypt.edu/shazeem/
[85] Handwritten Arabic Characters Database

				خط	۶،۵۳۲
[۲۱۷]	ACTIV	عربی	منظره (مصنوعی)	کادر متن[86]/ خط متن/ کلمه	/۴،۰۶۵ /۴،۸۲۴ ۲۱،۵۲۰
[۲۴۲]	APTI	عربی	چاپی (مصنوعی)	حرف/ زیر کلمه/ کلمه	/۲۵۹،۳۱۲،۰۰۰ /۱۰۹،۹۳۳،۲۰۰ ۴۵،۳۱۳،۶۰۰
[۱۸۱]	CENPARMI	عربی	دستنویس	تاریخ/ رقم/ رشته ای از ارقام/ حرف/ کلمه	/۲۸۴ /۴۶،۸۰۰ /۱۳،۴۳۹ /۲۱،۴۲۶ ۱۱،۳۷۵
[۲۴۳]	UPTI[87]	اردو	چاپی (مصنوعی)	خط/ لیگاتور	/- ۱۰،۰۰۰
[۱۸۳]	CENPARMI	اردو	دستنویس	تاریخ/ رقم/ رشته ای از ارقام/ حرف/ کلمه	/۳۱۸ /۶۰،۳۲۹ /۱۲،۹۱۴ /۱۴،۸۹۰ ۱۹،۴۳۲
[۲۴۴]	CENPARMI	دری	دستنویس	تاریخ/ حرف/ رقم/ کلمه/ حروف خاص	/۲۰۰ /۷۴۰۰ /۲۸،۰۰۰ /۱۴،۶۰۰ ۱۴۰۰

[86] Textbox
[87] Urdu Printed Text Images

۵. معیارهای ارزیابی

معیارهای ارزیابی سامانه های بازشناسی حروف به دو دسته کلی بازشناسی در سطح نویسه و بازشناسی در سطح کلمه تقسیم می شوند. معیار دقت در سطح نویسه، به صورت نسبت تعداد نویسه هایی که به درستی بازشناسی شده اند، به تعداد کل نویسه های موجود در متن صحیح به دست می آید. معیار دیگری به نام صحت[88] در سطح نویسه ها وجود دارد که به صورت تعداد نویسه هایی که به درستی بازشناسی شده اند، به تعداد کل نویسه های موجود در متن خروجی که توسط سامانه بازشناسی حروف تولید شده است، به دست می آید.

معیار دقیق تری که برای ارزیابی نحوه عملکرد سامانه های بازشناسی حروف در سطح نویسه وجود دارد، فاصله لونشتاین[89] است. این فاصله به صورت کم‌ترین تعداد درج، حذف و جایگزینی های لازم نویسه ها تعریف می شود تا هر دو متن با یکدیگر یکسان شوند. دقت بازشناسی بر اساس این فاصله با رابطه (۷) به دست می آید [۱۰۸].

$$\text{دقت} = 100 * (1 - \frac{\text{حذف} + \text{جایگزینی} + \text{درج}}{\text{طول کل متن صحیح}}) \quad (۷)$$

ارزیابی دقت بازشناسی در سطح کلمه به صورت نسبت تعداد کلمه هایی که تمام نویسه های آن به طور صحیح بازشناسی شده اند، به تعداد کل کلمه های موجود در متن اصلی است [۲۴۵]. روش های ارزیابی عملکرد در سطح کلمه، خود به دو دسته ارزیابی بر اساس کل کلمه ها یا ارزیابی بر اساس کلمه های غیر ایست-واژه تقسیم می شوند. در همه زبان ها، کلمه هایی از جمله از، با، به، را و ... هستند که در متون مختلف به وفور یافت می شوند، اما ارتباط معنایی خاصی با متن ندارند. به عبارت دیگر، حذف این کلمه ها، لطمه ای به کاربردهایی نظیر طبقه بندی متون نمی زند. یکی از روش های ارزیابی کارایی یک سامانه بازشناسی الگو، بررسی دقت آن روش در تشخیص صحیح کلمه های غیر ایست-واژه است. بدین منظور، یک جدول مراجعه[90] از ایست —واژه ها نیاز است تا ایست-واژه های موجود در متن نادیده گرفته شوند. سپس، دقت بازشناسی به صورت نسبت تعداد کلمه های غیر ایست-واژه ای که به درستی بازشناسی شده اند، به تعداد کل کلمه های غیر ایست-واژه موجود در متن صحیح به دست می آید [۲۴۶].

علاوه بر معیارهای ذکر شده، معیاری مختص خط فارسی و عربی به نام دقت حروف بر اساس کلاس حرف، موجود است که دقت بازشناسی را در سطح نویسه انجام می دهد. در این روش، هر نویسه فارسی یا عربی، بر مبنای جایگاه نویسه در کلمه (مستقل، ابتدایی، میانی و انتهایی) و همچنین بر اساس تعداد نقطه ها، در دسته های مختلفی جای می گیرند [۲۴۷].

۶. سامانه های تجاری/تحقیقاتی آزمایش شده بازشناسی نوری حروف

در ده های گذشته، نرم افزارهای مختلفی اعم از اختصاصی و متن باز[91]، برای بازشناسی نوری حروف توسعه یافته اند. از جمله معروف‌ترین نرم افزارهای اختصاصی، می‌توان به Abby FineReader، Readiris 16، Adobe Acrobat 11 و Omnipage اشاره کرد [۲۴۸]. نرم افزار Abby که معمولا به عنوان مرجعی برای مقایسه و ارزیابی روش های جدید استفاده می شود، از زبان های مختلف و انواع فرمت های ورودی/خروجی تصاویر پشتیبانی می کند. مشکل بزرگی که نرم افزار Abby دارد، این است که تصاویر را به صورت تکی دریافت می کند و بارگذاری تصاویر در یک پوشه و انجام بازشناسی به صورت یک جا برای تمام تصاویر، تنها در نسخه "Enterprise" آن ممکن است. Abby، همچنین، محصول خود را به صورت SDK، سرویس های برخط/ابری و نسخه مخصوص سرور ارائه می دهد. برخی از محصولات این شرکت، از لینوکس نیز پشتیبانی می کند. نرم افزار Readiris نیز از کارایی خوبی برای بازشناسی برخوردار است. این نرم افزار نیز مشابه Abby، کار روی تصاویر را یکی یکی انجام می‌دهد و بارگذاری تصاویر به صورت یک جا، تنها در نسخه Pro آن فراهم است. نسخه مربوط به تست نرم افزار، تنها روی ۱۰۰ تصویر جواب می دهد و بنابراین، شرایط آزمودن این نرم افزار محدود است. این نرم افزار از محیط گرافیکی خوبی برخوردار است و قابلیت استفاده روی بسترهای ویندوز، مک، iOS و اندروید را دارد. Adobe Acrobat مجموعه ای برای مدیریت و خواندن فایل های pdf است و نرم افزاری مختص بازشناسی نوری حروف نیست. این نرم افزار، هیچ ابزاری برای پیش پردازش دستی یا منطقه بندی تصاویر ندارد. نسخه رایگان بازشناسی نوری حروف، برای مدت چند روز روی این نرم افزار فعال می ماند. Omnipage نیز یکی دیگر از نرم افزارهای

[88] Precision
[89] Levenshtein
[90] Lookup table
[91] Open source

اختصاصی OCR است که قابلیت نصب و اجرا روی سیتم عامل های ویندوز، مک و لینوکس را دارد.

هیچ یک از بسته های متن باز OCR دارای محیط گرافیکی برای کاربر نیستند. همچنین، این بسته ها، از کیفیت کار پایین تری نسبت به نرم افزارهای اختصاصی غیر رایگان، برخوردارند. یکی از کارهایی که می توان برای افزایش کیفیت کار این بسته های متن باز انجام داد، اعمال مراحل پیش پردازش روی صفحات پویش شده است. بدین منظور، می توان از نرم افزاری نظیرScan Tailor برای افزایش کیفیت تصاویر پویش شده و در نتیجه افزایش دقت کار استفاده کرد. معروف ترین بسته متن باز و رایگان بازشناسی نوری حروف Tesseract نام دارد که در ابتدا به وسیله شرکت HP پایه گذاری شد و در حال حاضر، شرکت گوگل، توسعه آن را بر عهده دارد. این بسته، زبان های زیاد و نیز چند فرمت عکسی متفاوت را پشتیبانی می کند. این بسته می تواند خروجی با فرمت hOCR را تولید کند که شامل اطلاعات چیدمان صفحه[92]، مقدار اطمینان از نویسه ها، کادر محصور کننده نویسه ها[93] و اطلاعات سبک[94] است. این بسته که می تواند به عنوان یک کتابخانه در برنامه ها فرا خوانده شود، قابلیت آموزش برای زبان ها و سمبل های جدید را دارد. Tesseract علاوه بر اینکه تجزیه و تحلیل مربوط به چیدمان صفحه پویش شده را انجام می دهد، دارای روش های پیش پردازشی نظیر تعیین جهت متن و تصحیح انحراف های جزئی صفحه می باشد. از دیگر بسته های متن باز بازشناسی نوری حروف می توان به GNU Ocrad، Gocr، Ocropus و Cuneiform اشاره کرد. به روز رسانی این نسخه ها معمولا به ندرت انجام می شود.

به غیر از نرم افزارها و بسته های متن باز معرفی شده، Google docs نیز که اساسا یک فضای ابری برای ذخیر فایل ها است، یک سرویس بر خط بسیار قدرتمند برای بازشناسی نوری حروف فراهم کرده است. این سرویس، به طور خودکار، عمل بازشناسی را روی هر تصویر یا pdf بارگذاری شده انجام می دهد و پس از پایان کار، متن بازشناسی شده قابل بارگیری است. برای استفاده از این سرویس، داشتن حساب کاربری گوگل ضروری است.

۷. جمع بندی

موفقیت های روز افزون شبکه های عمیق در کاربردهای مختلف، توجه محققان در حوزه بازشناسی نوری حروف را نیز به خود جلب کرد؛ به طوری که اکثر قریب به اتفاق تحقیقاتی که در سال های اخیر در این حوزه مطرح شده است، مبتنی بر شبکه های عمیق هستند. در این تحقیق، روش های بازشناسی حروف مبتنی بر شبکه های ژرف مرور و کارایی آن ها در مقابل روش هایی که مبتنی بر شبکه های ژرف نیستند، مورد ارزیابی قرار گرفت. یکی از موفقیت های این شبکه ها در مسأله بازشناسی نوری حروف این است که معماری به کار برده شده، مستقل از نوع خط است و با استفاده از ساختار یکسان و تنها با آموزش شبکه، می توان از آن برای بازشناسی زبان های مختلف استفاده کرد. نکته حائز اهمیت دیگر این است که در ساختار دو روش غالب بازشناسی حروف مبتنی بر شبکه های ژرف (روش مبتنی بر CTC و روش های مبتنی بر ساز و کار توجه) نیازی به جداسازی حروف به نویسه ها نیست، بلکه این روش ها می توانند به طور مستقیم، کلمه ها را بازشناسی کنند. همچنین، هم تراز سازی متن صحیح با تصویر ورودی، در دو روش مبتنی بر CTC و ساز و کار توجه، انجام می شود و در نتیجه بر خلاف روش های مستقل از الگوریتم های ژرف، فرایند وقت گیر و هزینه بر هم تراز سازی، عملا حذف می شود. یکی دیگر از مزایایی که سامانه های بازشناسی حروف مبتنی بر الگوریتم های ژرف دارند، قابلیت تعمیم پذیری بالاتر آن ها نسبت به روش های مستقل از الگوریتم های ژرف، برای متن های دیده نشده است. همانطور که پیش از این اشاره شد، مرجع [۹۳]، برای اثبات قابلیت تعمیم پذیری روش پیشنهادی خود که بر پایه شبکه LSTM است، آموزش شبکه را با استفاده از داده های آموزشی مصنوعی انجام داد، اما کارایی شبکه را روی داده های آزمون واقعی، که شامل صفحات اسکن شده کتاب های مختلف بود، سنجید. در کنار دقت های بالایی که در بازشناسی حروف با استفاده از این شبکه ها به دست می آید، مدت زمان زیاد برای آموزش شبکه و همچنین لزوم و جود تعداد قابل توجه نمونه های آموزشی را نباید از یاد برد، به طوری که رسیدن به دقت مناسب، با تعداد نمونه های آموزشی ارتباط نزدیک دارد [۹۵]. برای کاهش زمان آموزش شبکه، یکی از راه های موجود، استفاده از روش انتقال یادگیری[95] است [۱۵۳]. جنبه منفی دیگری که در رابطه با روش های مبتنی بر الگوریتم های ژرف وجود دارد، این است که آن ها مدل های جعبه سیاه هستند؛ بدین معنی که اطلاعات بسیار ناچیزی از چرایی و چگونگی رسیدن به کارایی فوق العاده آن ها در دست است.

با توجه به مطالب مطرح شده در مقاله، دیدیم که روش های رایج برای بازشناسی حروف که مبتنی بر شبکه های ژرف هستند، به دو روش کلی مبتنی بر CTC و مبتنی بر ساز و کار توجه تقسیم می شوند. اخیرا، تحقیقاتی برای اعمال هر دوی

[92] Layout information
[93] Bounding box
[94] Style information
[95] Transfer learning

این روش ها انجام شده است که از دقت و سرعت مناسب نیز برخوردارند [۱۱۹-۱۲۱]. مرجع [۲۴۹]، این دو روش بازشناسی را روی مسأله بازشناسی جملات در مناظر با استفاده از مجموعه های داده واقعی و با ابعاد بالا مقایسه کرده است. آن ها بر اساس آزمایش های گسترده، توصیه های عملی برای محققان ارائه داده اند. به طور مثال، روش های مبتنی بر توجه، می توانند به دقت بالاتری روی کلمات جدا سازی شده در مقایسه با روش های مبتنی بر CTC دست یابند؛ در حالیکه کارایی ضعیف تری روی بازشناسی جملات، نسبت به روش های مبتنی بر CTC دارند. همچنین، در مواردی نظیر بازشناسی حروف در تصاویر مربوط به معادلات ریاضی که حروف و اعداد ممکن است به صورت کسری و زیر هم نوشته شوند، اساسا روش های مبتنی بر CTC دقت کافی را ندارند. در این موارد، روش های مبتنی بر توجه نتایج بسیار بهتری نسبت به CTC دارند [۲۵۰]. بنابراین، انتخاب روش بازشناسی باید با توجه به مسأله صورت گیرد. انجام تحقیقاتی برای یافتن سایر روش ها جهت رقابت و جایگزینی دو روش موجود نیز می تواند بسیار ارزشمند باشد. به طور مثال، تابع تراکم آنتروپی متقاطعی[96] که در مرجع [۱۲۲] برای جایگزینی CTC و روش های توجه، طراحی شده، علاوه بر آن که عملکرد رقابتی خوبی در مقابل این دو روش داشته است، پیاده سازی سریع تری دارد و حجم حافظه کمتری مورد نیاز است.

یک مزیت یکتای روش های توجه یا شبکه های رمزگذار-رمز گشا این است که یک مدل زبانی به راحتی می تواند در بالای رمزگشا ادغام شود، در حالیکه این کار در روش های مبتنی بر CTC به راحتی انجام نمی شود. مدل زبان خطوط متن ورودی و خروجی را به عنوان دنباله ای از کاراکترها (به جای کلمات) مشاهده می کند و پیش بینی هر نویسه، صراحتا به نویسه قبل از آن مشروط می شود. [۱۳۸].

یکی از زمینه های بالقوه ای که برای تحقیقات بیشتر وجود دارد، اصلاح ساختار CTC برای کار با متن های دارای اعوجاج های فضایی نظیر نوشته های موجود در منظره است. با توجه به بهبود نسبی که در مرجع [۱۲۸] برای مدل دو بعدی CTC نسبت به CTC استاندارد ارائه شده است، تحقیقات بیشتر در این زمینه می توانند ارزشمند باشند.

در کاربردهای واقعی، نیاز است تا زبان متون مورد نظر مشخص باشد تا بتوان بازشناسی حروف را انجام داد. با توجه به اینکه روش های مبتنی بر الگوریتم های ژرف بازشناسی حروف، مستقل از زبان هستند و یک ساختار شبکه ای یکسان را می توان روی زبان های مختلف، آموزش داد و از آن نتیجه گرفت، پیشنهاد می شود مجموعه های داده چند زبانه توسعه پیدا کنند که به مرحله تشخیص زبان نیاز نباشد. در حال حاضر، تعداد مجموعه های داده چند زبانی محدود است [۱۹۸, ۱۹۹] و تا آنجا که می دانیم، مجموعه داده چند زبانی فارسی-انگلیسی غنی اعم از متنون چاپی یا متن های موجود در منظره، وجود ندارد.

تشکر و قدردانی

نویسندگان بر خود لازم می دانند، مراتب تشکر صمیمانه را از آقای دکتر کبیر، سردبیر محترم مجله ماشین بینایی و پردازش تصویر ایران به پاس راهنمایی های ایشان و همچنین، داوران محترم، به پاس نظرات و پیشنهادات سازنده آن ها، که قطعا به بهبود خوانایی و کیفیت مقاله کمك کرده است، اعلام دارند.

[96] Aggregation cross-entropy function

مراجع

[1] A. Singh, K. Bacchuwar, and A. Bhasin, "A survey of OCR applications," *International Journal of Machine Learning and Computing,* vol. 2, no. 3, p. 314, 2012.

[2] 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, p. 1950001, 2019.

[3] S. Kashef, H. Nezamabadi-Pour, and E. Rashedi, "Adaptive enhancement and binarization techniques for degraded plate images," *Multimedia Tools and Applications,* vol. 77, no. 13, pp. 16579-16595, 2018.

[۴] س. رخشانی اول, ع. راشدی, and ح. نظام آبادی پور, "بازشناسی پلاک خودرو با استفاده از یادگیری ژرف," نشریه ماشین بینایی و پردازش تصویر، دوره ۶، شماره ۱، صفحات ۳۱-۴۶، ۱۳۹۸.

[5] 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.

[6] R. Gossweiler, M. Kamvar, and S. Baluja, "What's up CAPTCHA? A CAPTCHA based on image orientation," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 841-850 .

[7] S. S. Tsai, H. Chen, D. Chen, G. Schroth, R. Grzeszczuk, and B. Girod, "Mobile visual search on printed documents using text and low bit-rate features," in *2011 18th IEEE International Conference on Image Processing*, 2011: IEEE, pp. 2601-2604 .

[8] R. Chen, B. C. Desai, and C. Zhou, "CINDI robot: an intelligent Web crawler based on multi-level inspection," in *11th International* Database Engineering and Applications Symposium (IDEAS 2007), 2007: IEEE, pp. 93-101 .

[9] Y. K. Ham, M. S. Kang, H. K. Chung, R.-H. Park, and G. T. Park, "Recognition of raised characters for automatic classification of rubber tires," *Optical Engineering,* vol. 34, no. 1, pp. 102-110, 1995.

[10] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: A survey," *IEEE transactions on pattern analysis and machine intelligence,* vol. 24, no. 2, pp. 237-267, 2002.

[۱۱] ا. فرامرزی, "بازشناسی نوری حروف: مروری بر مباحث نظری و ملاحظات کاربردی با تاکید بر مسائل خاص زبان فارسی," پژوهشنامه پردازش و مدیریت اطلاعات، دوره ۲۰، شماره ۳ و ۴، صفحات ۳۳-۶۱، ۱۳۸۴.

[12] J. Sadri, M. R. Yeganehzad, and J. Saghi, "A novel comprehensive database for offline Persian handwriting recognition," *Pattern Recognition,* vol. 60, pp. 378-393, 2016.

[13] B. Parhami and M. Taraghi, "Automatic recognition of printed Farsi texts," *Pattern Recognition,* vol. 14, no. 1-6, pp. 395-403, 1981.

[14] H. Khosravi and E. Kabir, "A blackboard approach towards integrated Farsi OCR system," *International Journal of Document Analysis and Recognition* (*IJDAR*), vol. 12, no. 1, pp. 21-32, 2009.

[15] K. Fouladi, B. N. Araabi, and E. Kabir, "A fast and accurate contour-based method for writer-dependent offline handwritten Farsi/Arabic subwords recognition," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 17, no. 2, pp. 181-203, 2014.

[۱۶] و. قدس and ا. کبیر, "بررسی شیوه های متداول نگارش دست نوشته های برخط فارسی به منظور استفاده در بازشناسی آن ها," مجله مهندسی برق دانشگاه تبریز، دوره ۴۱، شماره ۱، پیاپی ۶۱، صفحات ۲۲-۳۲، ۱۳۹۰.

[۱۷] م. شکوهی and ح. شیدائیان, "مروری بر سیستم های OCR به منظور شناسایی متون فارسی-عربی-اردو," دومین کنفرانس بین المللی نوآوری در علوم کامپیوتر و مهندسی برق, اسفند ۱۳۹۷.

[18] S. N. Srihari, "Handwritten address interpretation: a task of many pattern recognition problems," *International journal of pattern recognition and artificial intelligence,* vol. 14, no. 05, pp. ۶۶۳-۶۷۴, ۲۰۰۰.

[19] S. Impedovo, P. S.-p. Wang, and H. Bunke, *Automatic bankcheck processing*. World Scientific, 1997.

[۲۰] س. رضوی and ا. کبیر, "بازشناسی برخط حروف مجزای فارسی با شبکه عصبی," سومین کنفرانس ماشین بینایی و پردازش تصویر ایران, اسفند ۱۳۸۳.

[21] R. Smith, "Limits on the application of frequency-based language models to OCR," in *2011 International Conference on Document Analysis and Recognition*, 2011: IEEE, pp. 538-542 .

[22] H. Al-Rashaideh, "Preprocessing phase for Arabic word handwritten recognition," *Information Process* (*Russian*), vol. 6, no. 1, 2006.

[23] F. Farooq, V. Govindaraju, and M. Perrone, "Pre-processing methods for handwritten Arabic documents," in *Eighth International Conference on Document Analysis and Recognition* (*ICDAR'05*) ۲۰۰۵: ,IEEE, pp. 267-271 .

[24] M. Dehghan, K. Faez, M. Ahmadi, and M. Shridhar, "Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM," *Pattern Recognition,* vol. 34, no. 5, pp. 1057-1065, 2001.

[۲۵] ع. عابدی and ا. کبیر, "فراتفکیک پذیری مبتنی بر نمونه تک تصویر متن با روش نزول گرادیان ناهمزمان ترتیبی," نشریه مهندسی برق و کامپیوتر ایران، دوره ۱۴، شماره ۳، صفحات ۱۷۷-۱۹۲، ۱۳۹۵.

[۲۶] م. فرکی and م. پالهنگ, "بازشناسی برخط حروف فارسی بر پایه مدل مخفی مارکوف," مجله مهندسی برق، دوره ۴۰، شماره ۱، پیاپی ۵۹، صفحات ۲۳-۳۴، ۱۳۸۹.

[27] H. A. Al Hamad, "Skew Detection/Correction and Local Minima/Maxima Techniques for Extracting a

New Arabic Benchmark Database," *International Journal of Advanced Computer Science and Applications* (*IJACSA*), vol ,۶ .no. 9, pp. 1-10, 2015.

[28] M. Ziaratban and K. Faez, "Non-uniform slant estimation and correction for Farsi/Arabic handwritten words," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 12, no. 4, p. 249, 2009.

[29] M. Ziaratban and K. Faez, "Detection and compensation of undesirable discontinuities within the farsi/arabic subwords," *Int. Arab J. Inf. Technol.*, vol. 8, no. 3, pp. 293-301, 2011.

[۳۰] ر. عزمی, ا. کبیر, and ک. بدیع, "بازشناسی حروف چاپی با استفاده از ویژگیهای منحنی پیرامونی," نشریه علوم و مهندسی کامپیوتر، دوره ۱، شماره ۱، صفحات ۲۹-۳۷، ۱۳۸۲.

[31] A. Lawgali, "A survey on Arabic character recognition," 2015.

[32] A. Chaudhuri, K. Mandaviya, P. Badelia, and S. K. Ghosh, "Optical character recognition systems," in *Optical Character Recognition Systems for Different Languages with Soft Computing*: Springer, 2017, pp. 9-41.

[33] K. Mohiuddin and J. Mao, "A comparative study of different classifiers for handprinted character recognition," in *Machine Intelligence and Pattern Recognition*, vol. 16: Elsevier, 1994, pp. 437-448.

[۳۴] م. میردارمنصورپناهی and ر. طاولی, "ارائه روش هوشمند برای تشخیص کلمات دستنویس فارسی با استفاده از ویژگی های آماری و دسته بند ماشین بردار پشتیبان," سومین کنفرانس بین المللی مهندسی کامپیوتر و سیستم های خبره, آبان ۱۳۹۵.

[35] H. Khosravi and E. Kabir, "Farsi font recognition based on Sobel–Roberts features," *Pattern Recognition Letters*, vol. 31, no. 1, pp. 75-82, 2010.

[36] A. Amin and J. F. Mari, "Machine recognition and correction of printed Arabic text," *IEEE Transactions on systems, man, and cybernetics*, vol. 19, no. 5, pp. 1300-1306, 1989.

[37] M. S. Khorsheed and W. F. Clocksin, "Structural Features of Cursive Arabic Script," in *BMVC*, 1999: Citeseer, pp. 1-10 .

[38] M. S. Khorsheed and W. F. Clocksin, "Multi-font Arabic word recognition using spectral features," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, 2000, vol. 4: IEEE, pp. 543-546 .

[39] J. H. AlKhateeb, J. Ren, J. Jiang, S. S. Ipson, and H. El Abed, "Word-based handwritten Arabic scripts recognition using DCT features and neural network classifier," in *2008 5th International Multi-Conference on Systems, Signals and Devices*, 2008: IEEE, pp. 1-5 .

[40] I. A. Jannoud, "Automatic Arabic hand written text recognition system," *American Journal of Applied Sciences*, vol. 4, no. 11, pp. 857-864, 2007.

[41] N. B. Amor and N. E. B. Amara, "Multifont Arabic Characters Recognition Using HoughTransform and HMM/ANN Classification," *Journal of multimedia*, vol. 1, no ,۲ .pp. 50-54, 2006.

[42] S. N. Nawaz, M. Sarfraz, A. Zidouri, and W. G. Al-Khatib, "An approach to offline Arabic character recognition using neural networks," in *10th IEEE International Conference on Electronics, Circuits and Systems, 2003. ICECS 2003 .Proceedings of the 2003*, 2003, vol. 3: IEEE, pp. 1328-1331 .

[43] S. Alma'adeed, "Recognition of off-line handwritten arabic words using neural network," in *Geometric Modeling and Imaging--New Trends* (*GMAI'06*), 2006: IEEE, pp. 141-144 .

[44] I. M. Khalaf Khatatneh and B. A.-R. El Emary, "Probabilistic Artificial Neural Network for Recognizing the Arabic. Hand Written Characters," in *Journal of Computer Science*, 2006: Citeseer .

[45] A. M. Asiri and M. S. Khorsheed, "Automatic Processing of Handwritten Arabic Forms using Neural Networks," in *IEC* (*Prague*), 2005, pp. 313-317 .

[46] A. Dehghani, F. Shabini, and P. Nava, "Off-line recognition of isolated Persian handwritten characters using multiple hidden Markov models," in *Proceedings International Conference on Information Technology: Coding and Computing*, 2001: IEEE, pp. 506-510 .

[47] S. Alma'adeed, C. Higgens, and D. Elliman, "Recognition of off-line handwritten Arabic words using hidden Markov model approach," in *Object recognition supported by user interaction for service robots*, 2002, vol. 3: IEEE, pp. 481-484 .

[48] H. A. Al-Muhtaseb, S. A. Mahmoud, and R. S. Qahwaji, "Recognition of off-line printed Arabic text using Hidden Markov Models," *Signal processing*, vol. 88, no. 12, pp. 2902-2912, 2008.

[49] J. H. AlKhateeb, F. Khelifi, J. Jiang, and S. S. Ipson, "A new approach for off-line handwritten Arabic word recognition using KNN classifier," in *2009 IEEE International Conference on Signal and Image Processing Applications*, pp. 191-194, 2009.

[۵۰] ا. ابراهیمی and ا. کبیر, "یک روش دو مرحله ای برای بازشناسی زیر-کلمات چاپی," نشریه مهندسی برق و کامپیوتر ایران،دوره ۲، شماره ۲، صفحات ۵۷-۶۲، ۱۳۸۳.

[۵۱] ح. خسروی and ا. کبیر, "بازشناسی متن چاپی فارسی بر مبنای جداسازی هوشمند," سومین کنفرانس بین المللی فناوری اطلاعات و دانش, ۱۳۸۶.

[52] L. Lorigo and V. Govindaraju, "Segmentation and pre-recognition of Arabic handwriting," in *Eighth International Conference on Document Analysis and Recognition* (*ICDAR'05*), 2005: IEEE, pp. 605-609 .

[53] J. Cowell and F. Hussain, "Thinning Arabic characters for feature extraction," in *Proceedings Fifth*

International Conference on Information Visualisation, 2001: IEEE, pp. 181-185 .

[54] T. Sari, L. Souici, and M. Sellami, "Off-line handwritten Arabic character segmentation algorithm: ACSA," in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002: IEEE, pp. 452-457 .

[55] R. Azmi and E. Kabir, "A new segmentation technique for omnifont Farsi text," *Pattern Recognition Letters,* vol ,۲۲ .no. 2, pp. 97-104, 2001.

[۵۶] ر. عزمی and ا. کبیر, "معرفی روش جدیدی برای جداسازی حروف در متون چاپی بدون توجه به نوع قلم," نشریه استقلال، دوره ۸، شماره ۲، صفحات ۱-۱۰، ۱۳۷۸.

[57] A. Hamid and R. Haraty, "A neuro-heuristic approach for segmenting handwritten Arabic text," in *Proceedings ACS/IEEE International Conference on Computer Systems and Applications*, 2001: IEEE, pp. 110-113 .

[58] D. Motawa, A. Amin, and R. Sabourin, "Segmentation of Arabic cursive script," in *Proceedings of the fourth international conference on document analysis and recognition*, 1997, vol. 2: IEEE, pp. 625-628 .

[59] S. Mozaffari and P. Bahar, "Farsi/Arabic handwritten from machine-printed words discrimination," in *2012 International Conference on Frontiers in Handwriting Recognition*, 2012: IEEE, pp. 698-703 .

[60] H. A. Yarmohammadi, A. A. Fard, and H. Khosravi, "Clustering low quality Farsi sub-words for word recognition," in *2014 Iranian Conference on Intelligent Systems* (*ICIS*), 2014: IEEE, pp. 1-5 .

[۶۱] ر. عزمی, ا. کبیر, and ک. بدیع, "ارائه یک الگوریتم دسته بندیبرای بازشناسی زیر کلمات چاپی," نشریه بین المللی مهندسی صنایع و مدیریت تولید، دوره ۱۲، شماره ۱، صفحات ۳۹-۴۹، ۱۳۸۰.

[62] A. Ebrahimi and E. Kabir, "A pictorial dictionary for printed Farsi subwords," *Pattern recognition letters,* vol. 29, no. 5, pp. 656-663, 2008.

[۶۳] ا. ابراهیمی, "استفاده از شکل کلی زیر کلمات چاپی در بازیابی تصویر مستندات و بازشناسی متون فارسی," رساله دکتری بخش مهندسی برق، دانشگاه تربیت مدرس تهران، ایران, ۱۳۸۴.

[۶۴] ح. خسروی and ا. کبیر, "ارزیابی روش های بازشناسی متون فارسی بر مبنای شکل کلی زیرکلمات," نشریه مهندسی برق و کامپیوتر ایران،دوره ۷، شماره ۴، صفحات ۲۶۷-۲۸۰، ۱۳۸۸.

[65] H. Davoudi, M. Cheriet, and E. Kabir, "Lexicon reduction of handwritten Arabic subwords based on the prominent shape regions," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 19, no. 2, pp. 139-153, 2016.

[66] H. Davoudi and E. Kabir, "Lexicon reduction for printed Farsi subwords using pictorial and textual dictionaries," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 17, no. 4, pp. 359-374, 2014.

[67] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, 1989.

[68] Z. Lu ,R. Schwartz, P. Natarajan, I. Bazzi, and J. Makhoul, "Advances in the bbn byblos ocr system," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99* (*Cat. No. PR00318*), 1999: IEEE, pp. 337-340 .

[69] M. R. Yousefi, M. R. Soheili, T. M. Breuel, and D. Stricker, "A comparison of 1D and 2D LSTM architectures for the recognition of handwritten Arabic," in *Document Recognition and Retrieval XXII*, 2015, vol. 9402: International Society for Optics and Photonics, p. ۹۴۰۲۰H .

[70] I. Bazzi, R. Schwartz, and J. Makhoul, "An omnifont open-vocabulary OCR system for English and Arabic," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 21, no. 6, pp. 495-504, 1999.

[71] Y. Kessentini, T. Paquet, and A. B .Hamadou, "Off-line handwritten word recognition using multi-stream hidden Markov models," *Pattern Recognition Letters,* vol. 31, no. 1, pp. 60-70, 2010.

[72] R. A.-H. Mohamad, L. Likforman-Sulem, and C. Mokbel, "Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 31, no. 7, pp. 1165-1177, 2008.

[73] A. Kundu, T. Hines, B. Huyck, J. Phillips, and L. C. Van Guilder, "Arabic handwriting recognition using variable duration HMM," in *Ninth International Conference on Document Analysis and Recognition* (*ICDAR 2007*), 2007, vol. 2: IEEE, pp. 644-648 .

[74] S. M. Touj, N. E. B. Amara, and H. Amiri, "A hybrid approach for off-line Arabic handwriting recognition based on a Planar Hidden Markov modeling," in *Ninth International Conference on Document Analysis and Recognition* (*ICDAR 2007*), 2007, vol. 2: IEEE, pp. 964-968 .

[۷۵] و. قدس and س. حسینی, "بازشناسی برخط زیر-کلمات فارسی بر اساس ویژگی‌های کدهای زنجیره‌ای فریمن با استفاده از مدل مخفی مارکوف," مجله ماشین بینایی و پردازش تصویر، دوره ۳، شماره ۱، پیاپی ۱، صفحات ۳۷-۴۴، ۱۳۹۵.

[76] M. Ashurpour and M. Ziaratban, "Online Handwritten Persian Isolated Letter Recognition by Using Discrete Markov Models and Language-Based Features," *Journal of Soft Computing and Information Technology,* vol. 6, no. 2, pp. 51-68, 2017.

[77] Z. Imani, Z. Ahmadyfard, and A. Zohrevand, "Holistic Farsi handwritten word recognition using gradient features," *Journal of AI and Data Mining* ,vol. 4, no. 1, pp. 19-25, 2016.

[78] B. Vaseghi, S. Alirezaee, M. Ahmadi, and R. Amirfattahi, "Off-line Farsi/Arabic handwritten word

international conference on Machine learning, 2006, pp. 369-376 .

[92] B. Su and S. Lu, "Accurate scene text recognition based on recurrent neural network," in *Asian Conference on Computer Vision*, 2014: Springer, pp. 35-48 .

[93] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, "High-performance OCR for printed English and Fraktur using LSTM networks," in *2013 12th International Conference on Document Analysis and Recognition*, 2013: IEEE, pp. 683-687 .

[94] R. W. Smith, "History of the Tesseract OCR engine: what worked and what didn't," in *Document Recognition and Retrieval XX*, 2013, vol. 8658: International Society for Optics and Photonics, p . ۸۶۵۸۰۲.

[95] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "Photoocr: Reading text in uncontrolled conditions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 785-792 .

[96] T. Wang, D. J. Wu, A. Coates, and A .Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings of the 21st international conference on pattern recognition* (*ICPR2012*), 2012: IEEE, pp. 3304-3308 .

[97] Y. Bengio, Y. LeCun, and D. Henderson, "Globally trained handwritten word recognizer using spatial representation, convolutional neural networks, and hidden Markov models," in *Advances in neural information processing systems*, 1994, pp. 937-944 .

[98] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *arXiv preprint arXiv:1406.2227,* 2014.

[99] C. Luo, Q. Lin, Y. Liu, L. Jin, and C. Shen, "Separating content from style using adversarial learning for recognizing text in the wild ",*arXiv preprint arXiv:2001.04189,* 2020.

[100] W. Wang *et al.*, "TextSR: Content-aware text super-resolution guided by recognition," *arXiv preprint arXiv:1909.07113,* 2019.

[101] W. Liu, C. Chen, and K.-Y. K. Wong, "Char-Net: A Character-Aware Neural Network for Distorted Scene Text Recognition," in *AAAI*, 2018, vol. 1, no. 2, p. 4 .

[102] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "Aster: An attentional scene text recognizer with flexible rectification," *IEEE transactions on pattern analysis and machine intelligence,* vol. 41, no. 9, pp. 2035-2048, 2018.

[103] W. Liu, C. Chen, K.-Y. K. Wong, Z. Su, and J. Han, "STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition," in *BMVC*, 2016, vol. 2, p. 7 .

recognition using vector quantization and hidden Markov model," in *2008 IEEE International Multitopic Conference*, 2008 :IEEE, pp. 575-578 .

[79] M. Dehghan, K. Faez, M. Ahmadi, and M. Shridhar, "Unconstrained Farsi handwritten word recognition using fuzzy vector quantization and hidden Markov models," *Pattern Recognition Letters,* vol. 22, no. 2, pp. 209-214, 2001.

[80] J .H. AlKhateeb, J. Ren, J. Jiang, and H. Al-Muhtaseb, "Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking," *Pattern Recognition Letters,* vol. 32, no. 8, pp. 1081-1088, 2011.

[81] Y. Bassil and M. Alwani, "Ocr post-processing error correction algorithm using google online spelling suggestion," *arXiv preprint arXiv:1204.0191,* 2012.

[82] Z. Khosrobeygi, H. Veisi, S. H. R. Ahmadi, and H. Shabanian, "A Rule-Based Post-Processing Approach to Improve Persian OCR Performance," *Scientia Iranica,* 2020.

[۸۳] ب. سمیه and مجید ایرانپور مبارکه, "بازشناسی کلمات دست نوشته با ویژگی های نوین و کاهش فرهنگ لغت," ماشین بینایی و پردازش تصویر، دوره ۴، شماره ۲، صفحات ۳۵-۴۷، ۱۳۹۶.

[84] H. Afli, L. Barrault, and H. Schwenk, "OCR Error Correction Using Statistical Machine Translation," *Int. J. Comput. Linguistics Appl.,* vol. 7, no. 1, pp. 175-191, 2016.

[۸۵] س. مسکنتی and ا. کشاورز, "تشخیص دست‌نوشتهٔ برخط فارسی با استفاده از مدل زبانی و کاهش قوانین نگارش کاربر," (in eng), فصل نامه علمی پردازش علائم و داده ها، دوره ۱۴، شماره ۲، صفحات ۳-۲۴، ۱۳۹۶.

[۸۶] پ. شیروانی, م. وطن-خواه-خوزانی, and خ. یغمایی, "بازشناسی متون فارسی با استفاده از مدل زبانی n-gram و پالایش گرامری," نشریه مهندسی برق و مهندسی کامپیوتر ایران، شماره ۱، پیاپی ۲۱، صفحات ۱۰۷-۱۱۵، ۱۳۹۳.

[87] Y. M. Alginahi, "A survey on Arabic character segmentation," *International Journal on Document Analysis and Recognition* (*IJDAR*), vol. 16, no. 2, pp. 105-126, 2013.

[88] R. G. Casey and E. Lecolinet, "A survey of methods and strategies in character segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 18, no. 7, pp. 690-706, 1996.

[89] Ø. D. Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition-a survey," *Pattern recognition,* vol. 29 ,no. 4, pp. 641-662, 1996.

[90] X. Chen, L. Jin, Y. Zhu, C. Luo, and T. Wang, "Text Recognition in the Wild: A Survey," *arXiv preprint arXiv:2005.03492,* 2020.

[91] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd*

networks," in *Advances in neural information processing systems*, 2009, pp. 545-552 .

[118] M. Yousef, K. F. Hussain, and U. S. Mohammed, "Accurate, data-efficient, unconstrained text recognition with convolutional neural networks," *Pattern Recognition,* vol. 108, p. 107482, 2020.

[119] W. Hu, X. Cai, J. Hou, S. Yi, and Z. Lin, "GTC: Guided Training of CTC towards Efficient and Accurate Scene Text Recognition," in *AAAI*, 2020, pp. 11005-11012 .

[120] L.-Q. Zuo, H.-M. Sun, Q.-C. Mao, R .Qi, and R.-S. Jia, "Natural scene text recognition based on encoder-decoder framework," *IEEE Access,* vol. 7, pp. 62616-62623, 2019.

[121] S. Kim, T. Hori, and S. Watanabe, "Joint CTC-attention based end-to-end speech recognition using multi-task learning," in *2017 IEEE international conference on acoustics, speech and signal processing* (*ICASSP*), 2017: IEEE, pp. 4835-4839 .

[122] Z. Xie, Y. Huang, Y. Zhu, L. Jin, Y. Liu, and L. Xie, "Aggregation cross-entropy for sequence recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6538-6547 .

[123] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013: IEEE, pp. 6645-6649 .

[124] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International conference on machine learning*, 2014, pp. 1764-1772 .

[125] P. He, W. Huang, Y. Qiao ,C. C. Loy, and X. Tang, "Reading scene text in deep convolutional sequences," *arXiv preprint arXiv:1506.04395,* 2015.

[126] Y. Gao, Y. Chen, J. Wang, M. Tang, and H. Lu, "Reading scene text with fully convolutional sequence modeling," *Neurocomputing,* vol ,٣٣٩ .pp. 161-170, 2019.

[127] X. Qi, Y. Chen, R. Xiao, C.-G. Li, Q. Zou, and S. Cui, "A Novel Joint Character Categorization and Localization Approach for Character-Level Scene Text Recognition," in *2019 International Conference on Document Analysis and Recognition Workshops* (*ICDAR W*), 2019, vol. 5: IEEE, pp. 83-90 .

[128] Z. Wan, F. Xie, Y. Liu, X. Bai, and C. Yao, "2D-CTC for scene text recognition," *arXiv preprint arXiv:1907.09705,* 2019.

[129] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078,* 2014.

[130] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances*

[104] B. Shi, X. Wang, P. Lyu ,C. Yao, and X. Bai, "Robust scene text recognition with automatic rectification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4168-4176 .

[105] M. Yang *et al.*, "Symmetry-constrained rectification network for scene text recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9147-9156 .

[106] F. Zhan and S. Lu, "Esir: End-to-end scene text recognition via iterative image rectification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2059-2068 .

[107] C. Luo, L. Jin, and Z. Sun, "Moran: A multi-object rectified attention network for scene text recognition," *Pattern Recognition,* vol. 90, pp. 109-118, 2019.

[108] A. Graves ,M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 31, no. 5, pp. 855-868, 2008.

[109] S. F .Rashid, F. Shafait, and T. M. Breuel, "Scanning neural network for text line recognition," in *2012 10th IAPR International Workshop on Document Analysis Systems*, 2012: IEEE, pp. 105-109 .

[110] M. Ziaratban and K. Faez, "A novel two-stage algorithm for baseline estimation and correction in Farsi and Arabic handwritten text line," in *2008 19th International Conference on Pattern Recognition*, 2008: IEEE, pp. 1-5 .

[111] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy ",*arXiv preprint arXiv:1710.10686,* 2017.

[112] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data,* vol. 6, no. 1, pp. 1-48, 2019.

[113] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 11, pp. 2298-2304, 2016.

[114] M. Namysl and I. Konya, "Efficient, lexicon-free OCR using deep learning," in *2019 International Conference on Document Analysis and Recognition* (*ICDAR*), 2019: IEEE, pp. 295-301 .

[115] S. Sarraf, "French Word Recognition through a Quick Survey on Recurrent Neural Networks Using Long-Short Term Memory RNN-LSTM," *arXiv preprint arXiv:1804.03683,* 2018.

[116] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[117] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural

[143] L. Kang, P. Riba, M. Rusiñol, A. Fornés, and M. Villegas, "Pay Attention to What You Read: Non-recurrent Handwritten Text-Line Recognition," *arXiv preprint arXiv:2005.13044,* 2020.

[144] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," *Advances in neural information processing systems,* vol. 27, pp. 2204-2212, 2014.

[145] J. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple object recognition with visual attention," *arXiv preprint arXiv:1412.7755,* 2014.

[146] C.-Y. Lee and S. Osindero, "Recursive recurrent nets with attention modeling for ocr in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2231-2239 .

[147] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou, "Aon: Towards arbitrarily-oriented text recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 201 ,٨pp. 5571-5579 .

[148] X. Yang, D. He, Z. Zhou, D. Kifer, and C. L. Giles, "Learning to Read Irregular Text with Attention Mechanisms," in *IJCAI*, 2017, vol. 1, no. 2, p. 3 .

[149] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: A simple and strong baseline for irregular text recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 8610-8617 .

[150] Y. Zhang, S. Nie, W. Liu, X. Xu, D. Zhang, and H. T. Shen, "Sequence-to-sequence domain adaptation network for robust text image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2740-2749 .

[151] Z. Wojna *et al.*, "Attention-based extraction of structured information from street view imagery ",in *2017 14th IAPR International Conference on Document Analysis and Recognition* (*ICDAR*), 2017, vol. 1: IEEE, pp. 844-850 .

[152] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," *Advances in neural information processing systems* ,vol. 28, pp. 2017-2025, 2015.

[153] M. A. KO and S. Poruran, "OCR-Nets: Variants of Pre-trained CNN for Urdu Handwritten Character Recognition via Transfer Learning," *Procedia Computer Science,* vol. 171, pp. 2294-2301, 2020.

[154] M. Elleuch, R. Maalej ,and M. Kherallah, "A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition," *Procedia Computer Science,* vol. 80, pp. 1712-1723, 2016.

[155] B. Alizadehashraf and S. Roohi, "Persian handwritten character recognition using convolutional neural network," in *2017 10th Iranian Conference on Machine Vision and Image Processing* (*MVIP*), 2017: IEEE, pp. 247-251 .

in neural information processing systems, 2014, pp . .٣١١٢-٣١٠٤

[131] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473,* 2014.

[132] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025,* 2015.

[133] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*, 2015, pp. 2048-2057 .

[134] Z. Yang ,D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480-1489 .

[135] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998-6008 .

[136] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805,* 2018.

[137] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020.

[138] J. Poulos and R. Valle" ,Character-based handwritten text transcription with attention networks," *arXiv preprint arXiv:1712.04046,* 2017.

[139] T. Bluche, J. Louradour, and R. Messina, "Scan, attend and read: End-to-end handwritten paragraph recognition with mdlstm attention ",in *2017 14th IAPR International Conference on Document Analysis and Recognition* (*ICDAR*), 2017, vol. 1: IEEE, pp. 1050-1055 .

[140] L. Kang, J. I. Toledo, P. Riba, M. Villegas, A. Fornés, and M. Rusinol, "Convolve, attend and spell: An attention-based sequence-to-sequence model for handwritten word recognition," in *German Conference on Pattern Recognition*, 2018: Springer, pp. 459-472 .

[141] J. Michael, R. Labahn, T. Grüning, and J. Zöllner, "Evaluating sequence-to-sequence models for handwritten text recognition," in *2019 International Conference on Document Analysis and Recognition* (*ICDAR*), 2019: IEEE, pp. 1286-1293 .

[142] T. Lupinski, A. Belaid, and A. K. Echi, "On the Use of Attention Mechanism in a Seq2Seq based Approach for Off-line Handwritten Digit String Recognition," in *2019 International Conference on Document Analysis and Recognition* (*ICDAR*), 2019: IEEE, pp. 502-507 .

[156] N. Javed, S. Shabbir, I. Siddiqi, and K. Khurshid, "Classification of Urdu ligatures using convolutional neural networks-a novel approach," in *2017 International Conference on Frontiers of Information Technology* (*FIT*), 2017: IEEE, pp. 93-97 .

[157] A. Zohrevand, M. Sattari, J. Sadri, Z. Imani, C. Y. Suen, and C. Djeddi, "Comparison of Persian Handwritten Digit Recognition in Three Color Modalities Using Deep Neural Networks," in *International Conference on Pattern Recognition and Artificial Intelligence*, 2020: Springer, pp. 125-136 .

[158] D. Ko, C. Lee, D. Han, H. Ohk, K. Kang, and S. Han, "Approach for Machine-Printed Arabic Character Recognition: the-state-of-the-art deep-learning method," *Electronic Imaging,* vol. 2018, no. 2, pp. 176-1-176-8, 2018.

[159] R. Maalej and M. Kherallah, "Convolutional neural network and BLSTM for offline Arabic handwriting recognition," in *2018 International Arab conference on information technology* (*ACIT*), 2018: IEEE, pp. 1-6 .

[160] M. Jain, M. Mathew, and C. Jawahar, "Unconstrained scene text and video text recognition for arabic script," in *2017 1st International Workshop on Arabic Script Analysis and Recognition* (*ASAR*), 2017: IEEE, pp. 26-30 .

[161] T. Anjum and N. Khan, "An attention based method for offline handwritten Urdu text recognition," in *2020 17th International Conference on Frontiers in Handwriting Recognition* (*ICFHR*), 2020: IEEE, pp. 169-174 .

[162] A. Mars and G. Antoniadis, "Arabic online handwriting recognition using neural network," *International Journal of Artificial Intelligence and Applications,* vol. 7, no. 5, pp. 51-59, 2016.

[163] A. Ashiquzzaman and A. K. Tushar, "Handwritten Arabic numeral recognition using deep learning neural networks," in *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition* (*icIVPR*), 2017: IEEE, pp. 1-4 .

[164] A. A. Alani, "Arabic handwritten digit recognition based on restricted Boltzmann machine and convolutional neural networks," *Information,* vol. 8, no. 4, p. 142, 2017.

[165] A. El-Sawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS Transactions on Computer Research,* vol. 5, pp. 11-19, 2017.

[166] R. Alaasam, B. Kurar, M. Kassis, and J. El-Sana, "Experiment study on utilizing convolutional neural networks to recognize historical Arabic handwritten text," in *2017١ st International Workshop on Arabic script analysis and recognition* (*ASAR*), 2017: IEEE, pp. 124-128 .

[167] M. Mudhsh and R. Almodfer, "Arabic handwritten alphanumeric character recognition using very deep neural network," *Information,* vol. 8, no. 3, p ,١٠٥ . .٢٠١٧

[168] K. S. Younis, "Arabic handwritten character recognition based on deep convolutional neural networks," *Jordanian Journal of Computers and Information Technology* (*JJCIT*), vol. 3, no. 3, pp. 186-200, 2017.

[169] G. Latif, J. Alghazo, L. Alzubaidi, M. M. Naseer, and Y. Alghazo, "Deep convolutional neural network for recognition of unified multi-language handwritten numerals," in *2018 IEEE 2nd International workshop on Arabic and derived script analysis and recognition* (*ASAR*), 2018: IEEE, pp .٩٥-٩٠ .

[170] L. Gui, X. Liang, X. Chang, and A. G. Hauptmann, "Adaptive Context-aware Reinforced Agent for Handwritten Text Recognition," in *BMVC*, 2018, vol. 207 .

[171] A. Ashiquzzaman, A. K. Tushar, A. Rahman, and F. Mohsin, "An efficient recognition method for handwritten arabic numerals using CNN with data augmentation and dropout," in *Data Management, Analytics and Innovation*: Springer, 2019, pp. 299-309.

[172] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Computing and Applications,* pp. 1-13, 2020.

[173] R. Hussain, A. Raza, I. Siddiqi, K. Khurshid, and C. Djeddi, "A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation," *EURASIP Journal on Image and Video Processing,* vol. 2015, no. 1, p. 46, 2015.

[174] U.-V. Marti and H. Bunke, "A full English sentence database for off-line handwriting recognition," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99* (*Cat. No. PR00318*), 1999: IEEE, pp. 705-708 .

[175] U.-V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition," *International Journal on Document Analysis and Recognition,* vol. 5, no. 1, pp. 39-4 ,ٶ .٢٠٠٢

[176] M. Liwicki and H. Bunke, "IAM-OnDB-an on-line English sentence database acquired from handwritten text on a whiteboard," in *Eighth International Conference on Document Analysis and Recognition* (*ICDAR'05*), 2005: IEEE, pp. 956-961 .

[177] E .Augustin, M. Carré, E. Grosicki, J.-M. Brodin, E. Geoffrois, and F. Prêteux, "RIMES evaluation campaign for handwritten mail processing," 2006 .

[178] R. Wilkinson *et al.*, "The first census optical character recognition systems conf.# NISTIR 4912," *The US*

[191] S. Al-Ma'adeed, D. Elliman, and C. A. Higgins" ,A data base for Arabic handwritten text recognition research," in *Proceedings eighth international workshop on frontiers in handwriting recognition*, 2002: IEEE, pp. 485-489 .

[192] N. Kharma, M. Ahmed, and R. Ward, "A new comprehensive database of handwritten Arabic words, numbers, and signatures used for OCR testing," in *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering* (*Cat. No. 99TH8411*), 1999, vol. 2: IEEE, pp. 766-768 .

[193] M. K. A .E. H. El and A. A. M. Alimi, "The On/Off (LMCA) Dual Arabic Handwriting Database".

[194] S. Mozaffari, H. El Abed, V. Märgner, K. Faez, and A. Amirshahi, "IfN/Farsi-Database: a database of Farsi handwritten city names," in *International Conference on Frontiers in Handwriting Recognition*, 2008 .

[195] H. Khosravi and E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties," *Pattern recognition letters,* vol. 28, no. 10, pp. 1133-1141, 2007.

[196] Y. Akbari ,M. J. Jalili, J. Sadri, K. Nouri, I. Siddiqi, and C. Djeddi, "A novel database for automatic processing of Persian handwritten bank checks," *Pattern Recognition,* vol. 74, pp. 253-265, 2018.

[197] U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals," *IEEE transactions on pattern analysis and machine intelligence,* vol. 31, no. 3, pp. 444-457, 2008.

[198] S. Al Maadeed, W. Ayouby, A. Hassaïne, and J. M. Aljaam, "Quwi: An arabic and english handwriting dataset for offline writer identification," in *2012 International Conference on Frontiers in Handwriting Recognition*, 2012: IEEE, pp. 746-751 .

[199] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "Cvl-database: An off-line database for writer retrieval, writer identification and word spotting," in *2013 12th international conference on document analysis and recognition*, 2013: IEEE, pp. 560-564 .

[200] G. Dimauro, S. Impedovo, R. Modugno, and G. Pirlo, "A new database for research on bank-check processing," in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002: IEEE, pp. 524-528 .

[201] E. Kavallieratou, N. Liolios, E. Koutsogeorgos, N. Fakotakis, and G. Kokkinakis, "The GRUHD database of Greek unconstrained handwriting," in *Proceedings of Sixth International Conference on Document Analysis and Recognition*, 2001: IEEE, pp. 561-565 .

Bureau of Census and the National Institute of Standards and Technology, Gaithersburg, MD, 1992.

[179] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11 , pp. 2278-2324, 1998.

[180] Y. Al-Ohali, M. Cheriet, and C. Suen, "Databases for recognition of handwritten Arabic cheques," *Pattern Recognition,* vol. 36, no. 1, pp. 111-121, 2003.

[181] H. Alamri, J. Sadri, C. Y. Suen, and N. Nobile, "A novel comprehensive database for Arabic off-line handwriting recognition," in *Proceedings of 11th International Conference on Frontiers in Handwriting Recognition, ICFHR*, 2008, vol. 8, pp. 664-669 .

[182] P. J. Haghighi, N. Nobile, C. L. He, and C. Y. Suen, "A new large-scale multi-purpose handwritten Farsi database," in *International Conference Image Analysis and Recognition*, 2009: Springer, pp. 278-286 .

[183] M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen, "A new large Urdu database for off-line handwriting recognition," in *International Conference on Image Analysis and Processing*, 2009: Springer, pp. 538-546 .

[184] M. Shah, C. He, N. Nobile, and C. Suen, "A handwritten Pashto database with multi-aspects for handwriting recognition," in *Proceedings of the 14th Conference of the International Graphonomics Society*, 2009, pp. 157-161 .

[185] F. Solimanpour, J. Sadri, and C. Y. Suen, "Standard databases for recognition of handwritten digits, numerical strings, legal amounts, letters and dates in Farsi language," 2 .٠٠۶

[186] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 16, no. 5, pp. 550-554, 1994.

[187] T. SAITO, "On the data base ETK9B of handprinted characters in JIS Chinese characters and its analysis," *IEICE trans,* vol. 68, no. 4, pp. 757-772, 1985.

[188] D.-H. KIM, Y.-S. Hwang, S.-T. Park, E.-J. Kim, S.-H. Paek, and S.-Y. BANG, "Handwritten Korean character image database PE92," *IEICE transactions on information and systems,* vol. 79, no. 7, pp. 943-950, 1996.

[189] T. Su, T. Zhang, and D. Guan, "Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text," *International Journal of Document Analysis and Recognition* (*IJDAR*), vol. 10 , no. 1, p. 27, 2007.

[190] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, "IFN/ENIT-database of handwritten Arabic words," in *Proc. of CIFED*, 2002, vol. 2: Citeseer, pp. 127-136 .

[214] F. Zhan, S. Lu, and C. Xue, "Verisimilar image synthesis for accurate detection and recognition of texts in scenes," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 249-٢٦٦.

[215] S. Long and C. Yao, "UnrealText: Synthesizing realistic scene text images from the unreal world," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5488-5497.

[216] S. Yousfi, S.-A. Berrani, and C. Garcia, "ALIF: A dataset for Arabic embedded text recognition in TV broadcast," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 1221-1225.

[217] O. Zayene, J. Hennebert, S. M. Touj, R. Ingold, and N. E. B. Amara, "A dataset for Arabic text detection, tracking and recognition in news videos-AcTiV," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 996-1000.

[218] A. Mishra, K. Alahari, and C. Jawahar, "Scene text recognition using higher order language priors," 2012.

[219] K. Wang and S. Belongie, "Word spotting in the wild," in *European Conference on Computer Vision*, 2010: Springer, pp. 591-604.

[220] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 robust reading competitions," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, 2003: Citeseer, pp. 682-687.

[221] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition challenge 2: Reading text in scene images," in *2011 international conference on document analysis and recognition*, 2011: IEEE, pp. 1491-1496.

[222] D. Karatzas *et al.*, "ICDAR 2013 robust reading competition," in *2013 12th International Conference on Document Analysis and Recognition*, 2013: IEEE, pp. 1484-1493.

[223] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.

[224] M. Tounsi, I. Moalla, and A. M. Alimi, "ARASTI: A database for Arabic scene text recognition," in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 2017: IEEE, pp. 140-144.

[225] M. Tounsi, I. Moalla, A. M. Alimi, and F. Lebouregois, "Arabic characters recognition in natural scenes using sparse coding for feature representations," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 1036-1040.

[202] D. Llorens *et al.*, "The UJIpenchars Database: a Pen-Based Database of Isolated Handwritten Characters," in *LREC*, 2008.

[203] M. Nakagawa and K. Matsumoto, "Collection of on-line handwritten Japanese character pattern databases and their analyses," *Document Analysis and Recognition,* vol. 7, no. 1, pp. 69-81, 2004.

[204] S. Gaur, S. Sonkar, and P. P. Roy, "Generation of synthetic training data for handwritten Indic script recognition," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 491-495.

[205] P. P. Roy, A. Mohta, and B. B. Chaudhuri, "Synthetic data generation for Indic handwritten text recognition," *arXiv preprint arXiv:1804.06254,* 2018.

[206] P. Krishnan and C. Jawahar, "Generating synthetic data for text recognition," *arXiv preprint arXiv:1608.04224,* 2016.

[207] R. Smith, "An overview of the Tesseract OCR engine," in *Ninth international conference on document analysis and recognition (ICDAR 2007)*, 2007, vol. 2: IEEE, pp. 629-633.

[208] F. K. Jaiem, S. Kanoun, M. Khemakhem, H. El Abed, and J. Kardoun, "Database for arabic printed text recognition research," in *International Conference on Image Analysis and Processing*, 2013: Springer, pp. 251-259.

[209] A. G. Al-Hashim and S. A. Mahmoud, "Printed Arabic text database (PATDB) for research and benchmarking," in *Proceedings of the 9th WSEAS international conference on Applications of computer engineering, ACE*, 2010, vol. 10, pp. 62-68.

[210] F. Chabchoub, Y. Kessentini, S. Kanoun, V. Eglin, and F. Lebourgeois, "SmartATID: A mobile captured Arabic Text Images Dataset for multi-purpose recognition tasks," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016: IEEE, pp. 120-125.

[211] I. Chtourou, A. C. Rouhou, F. K. Jaiem, and S. Kanoun, "ALTID: Arabic/Latin text images database for recognition research," in *2١٣ ٠١٥th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 836-840.

[212] F. Slimane, R. Ingold, S. Kanoun, A. M. Alimi, and J. Hennebert, "A new arabic printed text image database and evaluation protocols," in *2009١٠ th International Conference on Document Analysis and Recognition*, 2009: IEEE, pp. 946-950.

[213] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2315-2324.

[239] E. A. El-Sherif and S. Abdelazeem, "A Two-Stage System for Arabic Handwritten Digit Recognition Tested on a New Large Database," in *Artificial intelligence and pattern recognition*, 2007, pp. 237-242 .

[240] A. Lawgali, M. Angelova, and A. Bouridane, "HACDB: Handwritten Arabic characters database for automatic character recognition ",in *European Workshop on Visual Information Processing (EUVIP)*, 2013: IEEE, pp. 255-259 .

[241] S. A. Mahmoud *et al.*, "KHATT: An open Arabic offline handwritten text database," *Pattern Recognition,* vol. 47, no. 3, pp. 1096-1112, 2014.

[242] F. Slimane, R .Ingold, S. Kanoun, A. M. Alimi, and J. Hennebert, "Database and evaluation protocols for arabic printed text recognition," *DIUF-University of Fribourg-Switzerland,* 2009.

[243] N. Sabbour and F. Shafait, "A segmentation-free approach to Arabic and Urdu OCR," in *Document Recognition and Retrieval XX*, 2013, vol. 8658: International Society for Optics and Photonics, p. 86580N .

[244] M. I. Shah, J. Sadri, C. Y. Suen, and N. Nobile, "A new multipurpose comprehensive database for handwritten Dari recognition ",in *Eleventh International Conference on Frontiers in Handwriting Recognition (ICFHR), Montreal Canada*, 2008, pp. 635-640 .

[245] N. Dershowitz and A. Rosenberg, "Arabic character recognition," in *Language, culture, computation. Computing-theory and technology*: Springer, 2014, pp. 584-602.

[246] S. V. Rice, J. Kanai, and T. A. Nartker, "An evaluation of OCR accuracy," *Information Science Research Institute, 1993 Annual Research Report,* vol. 9, p. 20, 1993.

[247] S. Saber, A. Ahmed, and M. Hadhoud, "Robust metrics for evaluating Arabic OCR systems," in *International Image Processing, Applications and Systems Conference*, 2014: IEEE, pp. 1-6 .

[248] M. Tomaschek, "Evaluation of off-the-shelf OCR technologies," PhD Thesis, Masaryk University, 2018 .

[249] F. Cong, W. Hu, Q. Huo, and L. Guo, "A comparative study of attention-based encoder-decoder approaches to natural scene text recognition," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019: IEEE, pp. 916-921 .

[250] Y. Deng, A. Kanervisto, J. Ling, and A. M. Rush, "Image-to-markup generation with coarse-to-fine attention," in *International Conference on Machine Learning*, 2017: PMLR, pp. 980-989 .

[226] T. Quy Phan, P. Shivakumara, S. Tian, and C .Lim Tan, "Recognizing text with perspective distortion in natural scenes," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 569-576 .

[227] A. Risnumawan, P. Shivakumara, C. S. Chan, and C. L. Tan, "A robust arbitrary text detection system for natural scene images," *Expert Systems with Applications,* vol. 41, no. 18, pp. 8027-8048, 2014.

[228] D. Karatzas *et al.*, "ICDAR 2015 competition on robust reading," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 1156-1160 .

[229] A. Veit, T. Matera, L. Neumann, J. Matas, and S. Belongie, "Coco-text: Dataset and benchmark for text detection and recognition in natural images," *arXiv preprint arXiv:1601.07140,* 2016.

[230] C.-K. Ch'ng, C. S. Chan, and C.-L. Liu, "Total-text: toward orientation robustness in scene text detection," *International Journal on Document Analysis and Recognition (IJDAR),* vol. 23, no. 1, pp. 31-52, 2020.

[231] S. B. Ahmed, S. Naz, M. I. Razzak, and R. B. Yusof, "A novel dataset for English-Arabic scene text recognition (EASTR)-42K and its evaluation using invariant feature extraction on detected extremal regions," *IEEE access,* vol. 7, pp. 19801-19820, 2019.

[232] B. Shi *et al.*, "ICDAR2017 competition on reading chinese text in the wild (RCTW-17)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1: IEEE, pp. 1429-1434 .

[233] M. He *et al.*, "ICPR2018 contest on robust reading for multi-type web images," in۲۴ ۲۰۱۸ *th International Conference on Pattern Recognition (ICPR)*, 2018: IEEE, pp. 7-12 .

[234] L. Yuliang, J. Lianwen, Z. Shuaitao, and Z. Sheng, "Detecting curve text in the wild: New dataset and new solution," *arXiv preprint arXiv:1712.02170,* 2017.

[235] N. Nayef *et al.*, "ICDAR2019 robust reading challenge on multi-lingual scene text detection and recognition—RRC-MLT-2019," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019: IEEE, pp. 1582-1587 .

[236] "https://www.kaggle.com/amir137825/persianocrdataset/version/2".

[237] M. Ziaratban, K. Faez, and F. Bagheri, "FHT: An unconstraint Farsi handwritten text database," in *2009 10th International Conference on Document Analysis and Recognition*, 2009: IEEE, pp. 281-285 .

[238] S. Mozaffari, K. Faez, F. Faradji, M. Ziaratban, and S. M. Golzan, "A comprehensive isolated Farsi/Arabic character database for handwritten OCR research," 2006 .

شیما کاشف مدرک کارشناسی، کارشناسی ارشد و دکتری خود را به ترتیب در سال‌های ۱۳۹۰، ۱۳۹۲ و ۱۳۹۷ در رشته مهندسی برق-الکترونیک از دانشگاه شهید باهنر کرمان دریافت کرد. ایشان هم چنین، دوره یک سال و نیم پسا دکتری خود را در زمینه داده‌کاوی در دانشگاه مذکور و تحت راهنمایی آقای دکتر حسین نظام آبادی‌پور در سال ۱۳۹۸ به اتمام رساندند و هم‌اکنون به عنوان پژوهشگر ارشد در آزمایشگاه پردازش هوشمند داده مشغول به فعالیت هستند. دکتر کاشف، نویسنده و هم-نویسنده ده‌ها مقاله در ژورنال‌ها و کنفرانس‌های علمی است. زمینه‌های علاقه‌مندی ایشان، داده‌کاوی، پردازش‌تصویر، بازشناسی‌الگو، یادگیری ماشینی و پردازش زبان طبیعی است.

حسین نظام‌آبادی‌پور 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 ژورنال ها و کنفرانس‌های علمی بوده است. زمینه های علاقه‌مندی ایشان، شامل پردازش‌تصویر، بازشناسی‌الگو، رایانش نرم و الگوریتم‌های فراابتکاری است.

الهام شعبانی‌نیا در سال ۱۳۸۵ مدرک کارشناسی خود را در رشته مهندسی کامپیوتر از دانشگاه شهید باهنر کرمان دریافت کرد. سپس، مدرک کارشناسی ارشد و دکتری را در رشته مهندسی کامپیوتر به ترتیب از دانشگاه صنعتی شریف و دانشگاه اصفهان در سال های ۱۳۸۸ و ۱۳۹۷ دریافت نمود. ایشان در سال ۱۳۹۹ به عنوان استادیار به بخش مهندسی کامپیوتر دانشگاه صنعتی سیرجان پیوست. دکتر شعبانی‌نیا نویسنده و هم-نویسنده دهها مقاله (در ژورنال‌ها و کنفرانس‌های علمی) و کتب تخصصی بوده است. زمینه‌های علاقه‌مندی ایشان، شامل پردازش‌تصویر، بینایی‌ماشین و یادگیری عمیق است.