Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 13, Number 3, Autumn 2024, pp. 31-45
Received: 05/08/2024, Revised: 12/12/2024, Accepted: 12/23/2024

GCOCR: A Method for Recognition of Characters Using Gated Convolutional and Recurrent Networks

Seyed Mohammadreza Mousavi[1], Amir Bavafa Toosi[2*]
1- Faculty of computer engineering and information technology, Sadjad University, Mashhad, Iran.
2*- Faculty of computer engineering and information technology, Sadjad University, Mashhad, Iran.
mo.mousavi271@sadjad.ac.ir[1], abavafat@sadjad.ac.ir[2*]

Corresponding author's address: Amir Bavafa Toosi, Faculty of computer engineering and information technology, Sadjad University, Mashhad, Iran.

Abstract- Handwritten Text Recognition (HTR) remains a challenging problem due to the inherent variability in handwriting styles. Variations in stroke thickness, slant, character spacing, and letter formation introduce significant complexity in accurately interpreting handwritten text. These factors, combined with the diversity of individual writing habits, make reliable recognition particularly difficult. Traditional methods often struggle with these challenges, particularly when processing degraded or non-standardized input. This research introduces a novel Gated Convolutional Optical Character Recognition (GCOCR) architecture, which integrates an advanced preprocessing step—combining padding, desloping, deslanting, and data augmentation techniques—with gated convolutional neural networks and a bidirectional recurrent neural network (BiRNN). This holistic design significantly enhances the accuracy and robustness of Handwritten Text Recognition (HTR) systems by optimizing input quality prior to network processing, allowing for more precise feature extraction and reducing the influence of noise and irrelevant patterns. The proposed model is rigorously evaluated on the IAM dataset, a benchmark in the HTR field. The experimental results demonstrate that GCOCR achieves a character error rate (CER) of 4.14%, representing a substantial improvement of 3.7% over the current state-of-the-art methods on the IAM dataset.

Keywords- Convolutional Recurrent Neural Networks, gated convolution neural network, data augmentation, deep learning, handwriting recognition, Gated Convolutional Intelligent Character Recognition.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۳، شماره ۳، پائیز ۱۴۰۳، صفحه ۳۱-۴۵
دریافت: ۱۴۰۳/۰۲/۱۹، بازنگری: ۱۴۰۳/۰۹/۲۲، پذیرش: ۱۴۰۳/۱۰/۰۳

GCOCR: روشی برای تشخیص حروف براساس شبکه‌های دروازه‌دار کانولوشنی و بازگشتی

سید محمدرضا موسوی[۱]، امیر باوفای طوسی*[۲]
۱- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی سجاد، مشهد، ایران.
*۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی سجاد، مشهد، ایران.
[1]mo.mousavi271@sadjad.ac.ir, [2*]abavafat@sadjad.ac.ir

* نشانی نویسنده مسئول: امیر باوفا طوسی، مشهد، خیابان جلال آل احمد، دانشگاه صنعتی سجاد، دانشکده مهندسی برق و کامپیوتر.

چکیده- **تشخیص متن دست‌نویس (HTR) همچنان به دلیل تنوع ذاتی در سبک‌های دست‌نویسی یک مسئله چالش‌برانگیز است. تفاوت‌ها در ضخامت خطوط، زاویه نوشتار، فاصله بین حروف و شکل کلی حروف، پیچیدگی‌های زیادی را در تفسیر دقیق متن‌های دست‌نویس ایجاد می‌کند. این عوامل، همراه با تنوع در عادات نوشتاری افراد، تشخیص صحیح را به‌ویژه دشوار می‌سازد. روش‌های سنتی اغلب در مواجهه با این چالش‌ها، به‌ویژه در پردازش ورودی‌های تخریب‌شده یا غیر استاندارد، با مشکل مواجه می‌شوند. این پژوهش یک معماری جدید به نام تشخیص نوری کاراکتر با استفاده از شبکه‌های کانولوشنی گیت‌دار (GCOCR) معرفی می‌کند که یک مرحله پیش‌پردازش پیشرفته شامل ترکیب تکنیک‌های پدینگ، حذف شیب، انحنا و افزایش داده‌ها را با شبکه‌های عصبی کانولوشنی گیت‌دار و شبکه عصبی بازگشتی دوجهته (BiRNN) ادغام می‌کند. این طراحی جامع، دقت و استحکام سیستم‌های تشخیص متن دست‌نویس (HTR) را با بهینه‌سازی کیفیت ورودی پیش از پردازش شبکه، به‌طور قابل توجهی افزایش می‌دهد و امکان استخراج ویژگی‌های دقیق‌تر و کاهش تأثیر نویز و الگوهای نامربوط را فراهم می‌کند. مدل پیشنهادی به‌طور دقیق بر روی مجموعه داده IAM، که یک معیار در حوزه HTR است، ارزیابی شده است. نتایج تجربی نشان می‌دهد که GCOCR به نرخ خطای کاراکتر (CER) ۴/۱۴٪ دست یافته که نشان‌دهنده بهبود قابل توجه ۳/۷٪ نسبت به روش‌های پیشرفته فعلی در مجموعه داده IAM است.**

واژه‌های کلیدی: شبکه‌های عصبی بازگشتی کانولوشنی[۱]، شبکه‌های عصبی کانولوشنی دروازه‌دار[۲]، افزونگی داده[۳]، یادگیری عمیق، شناسایی دستخط[۴].

۱- مقدمه

شناخت کاراکترها، حروف یا اعداد برای انسان کار دشواری نیست، اما انجام همین عمل به‌وسیله ماشین کار نسبتاً پیچیده‌ای هست؛ زیرا انسان به بی‌نهایت روش ممکن قادر خواهد بود یک متن را انشاء کند [۱]. ازاین‌رو تشخیص دستخط یکی از پیچیده‌ترین زمینه‌های تحقیقاتی در حوزه پردازش تصویر و شناسایی الگو در سال‌های اخیر محسوب می‌شود. در چند دهه گذشته شناسایی الگوهای نوشتاری شامل علائم و نمادهای متداول در اسناد، توسط گروه‌های محققین؛ موردبررسی قرارگرفته است [۲], [۳], [۴]. برخی محققین برای گونه‌های تشخیص حروف تعاریف متفاوتی را متصورند. به‌عنوان‌مثال از عبارت الف- "شناسایی حروف" برای تشخیص علائم تولیدشده توسط کامپیوتر و برای تشخیص حروف دست‌نویس از عبارت ب- "شناسایی هوشمند حروف[۵]" استفاده کرده‌اند. با این رویکرد، تشخیص متن دست‌نوشته فعالیت ماشین جهت ادراک و ترجمه علائم دست‌نوشته منطقی، از متون اسکن‌شده و یا سایر چنین تجهیزات می‌باشد که بر اساس چگونگی به دست آمدن داده‌های ورودی به دسته‌های شناسایی دستخط برخط[۶] و برون‌خط[۷] دسته‌بندی شده‌اند که عمده‌ترین چالش آن‌ها، تنوع الگوهای دستخط نویسنده‌های مختلف می‌باشد [۵], [۶], [۷]. در حالت اول

(برخط) حرکت قلم توسط حسگر تفسیر می‌شود و شامل نمایش دیجیتالی تغییرات تکانه‌ای آن است که اطلاعات شیب[۸] و انحنای[۹] دست‌نوشته را توصیف می‌کنند. در حالت دوم (برون‌خط) نسخه‌ای از یک سند فیزیکی اسکن می‌شود که شامل ترجمه کاراکترهای دست‌نوشته از اسناد کپی شده است[۸]. معمولاً در حالت دوم قبل از پردازش بر روی این نسخه، از الگوریتم‌های رفع شیب و انحنا استفاده می‌شود.

در ابتدا سیستم‌های شناسایی دست‌نوشته از مدل‌های مخفی مارکوف[۱۰][۹]، [۱۰] بهره می‌بردند و با پیشرفت‌های بعدی شبکه‌های عصبی کانولوشنی [۱۱]، [۱۲]، [۱۳] ازجمله شبکه‌های عصبی کانولوشنی دروازه‌دار و شبکه‌های عصبی بازگشتی [۱۴]، [۱۵]، [۱۶] به‌صورت جداگانه و ترکیبی، تغییرات حائز اهمیتی را در این حوزه ایجاد کرده‌اند.

شبکه‌های عصبی کانولوشنی نتایج خوبی را در حوزه بازشناسی الگو و به‌خصوص شناسایی دست‌نوشته و قطعه‌بندی آن‌ها [۱۷]، [۱۸] از خود نشان داده‌اند. این شبکه‌ها با اعمال روش‌های افزونگی داده مانند اعوجاجات کشسان[۱۱][۱۹]، [۲۰] قادر خواهد بود تا کارایی را به مقدار قابل‌توجهی بهبود دهد. شبکه‌های عصبی کانولوشنی با هردو مشکل موجود یعنی طبقه‌بندی و تشخیص مقابله کرده و از جامعیت کلی برخوردار می‌باشند [۲۱]؛ به این معنا که این شبکه‌ها قادر به محاسبه توابع پیوسته با دقت مناسب هستند. شبکه‌های عصبی کانولوشنی دروازه‌دار نیز مشتقی از شبکه‌های کانولوشنی می‌باشند با این خاصیت که گرادیان‌ها می‌توانند در بین لایه‌ها انتشار یابند. به‌طور خلاصه، مکانیسم دروازه‌ای، مسیری که داده‌ها وارد شبکه می‌شوند را کنترل کرده و برای شبکه‌های عصبی بازگشتی مناسب می‌باشد. این مکانیسم تا حدی شبیه به شبکه‌های حافظه طولانی کوتاه‌مدت[۱۲] می‌باشد. شبکه‌های حافظه طولانی کوتاه‌مدت، حافظه درازمدتشان را از طریق یک سلول جداگانه فعال می‌کنند، که توسط ورودی و دروازه فراموشی[۱۳] کنترل می‌شوند. این مکانیسم باعث می‌شود که اطلاعات از طریق گام‌های زمانی[۱۴] زیادی جریان یابد. بدون این دروازه‌ها، اطلاعات می‌توانند به‌راحتی از طریق تغییراتی که در هر گام زمانی رخ می‌دهد محو شوند. این در حالی است که، شبکه‌های کانولوشنی از چنین محوشدگی گرادیان برخوردار نیستند و به دروازه‌های فراموشی نیاز ندارند [۲۲]. واحدهای خطی دروازه‌ای[۱۵] یک مکانیسم دروازه‌ای ساده‌شده بر اساس کار [۲۳] می‌باشند که برای دروازه‌های غیرقطعی[۱۶] طرح‌شده‌اند و مشکل محوشدگی گرادیان را با داشتن واحدهای خطی که با دروازه‌ها جفت شده‌اند را کاهش می‌دهد. این امر به گرادیان اجازه می‌دهد تا در طول واحدهای خطی، بدون مقیاس‌گذاری[۱۷] انتشار یابد و قابلیت‌های غیرخطی لایه‌ها را حفظ کند.

شبکه‌های عصبی بازگشتی، مانند شبکه حافظه طولانی کوتاه‌مدت یک تصویر ورودی را قطعه‌بندی کرده و به‌صورت یک دنباله آن را در نظر می‌گیرد [۲۴]. در سال‌های گذشته، ترکیب شبکه‌های کانولوشنی و بازگشتی پیشرفت قابل‌توجهی را برای شناسایی دست‌نوشته ارائه نموده است به‌عنوان‌مثال، چن و همکاران [۲۵]، از ترکیب شبکه‌های عصبی کانولوشنی و بازگشتی برای شناسایی زبان‌های مختلف بهره بردند. مشتقات شبکه‌های عصبی کانولوشنی مانند شبکه‌های عصبی کانولوشنی دروازه‌دار [۲۶] نیز بسیار موثر واقع‌شده‌اند. همچنین در تحقیق [۲۷]، تنها از شبکه‌های کانولوشنی کلاسیک استفاده‌شده است و نیاز به شبکه‌های بازگشتی رفع شده است.

علی‌رغم نتایج بی‌نظیر پژوهش‌های انجام‌گرفته در این حوزه، همواره محققان به دنبال ارائه مدل‌های جدیدتر با دقت مناسب و تعداد پارامترهای کم‌تر می‌باشند. ما در این تحقیق، دو روش پیشنهاد می‌کنیم. در روش اول یک معماری که تنها با استفاده از شبکه‌های عصبی کانولوشنی عمل شناسائی حروف را انجام می‌دهد ارائه می‌گردد، این معماری منجر به کاهش پارامترهای شبکه و بهبود نسبی دقت می‌شود. برای افزایش دقت، روش دومی ارائه می‌گردد که با استفاده از شبکه‌های بازگشتی و دروازه‌دار دقت افزایش می‌یابد. در معماری روش دوم ابتدا بر روی داده‌های ورودی پیش‌پردازش[۱۸] انجام می‌گیرد. در قسمت پیش‌پردازش، پس از اعمال لایه‌گذاری[۱۹] بر روی بلوک کلمات؛ شیب و انحنای حروف گرفته می‌شود. به جهت جلوگیری از بیش‌برازش داده‌ها، بر روی آن‌ها افزونگی داده اعمال می‌شود. در ادامه، داده‌ها برای ارسال به شبکه تعبیه شده شامل لایه‌های کانولوشنی دروازه‌دار و بازگشتی آماده می‌شوند.

این تحقیق بر توسعه یک معماری جدید تحت عنوان GCOCR تمرکز دارد که مرحله پیش‌پردازش پیشرفته‌ای را شامل تکنیک‌های پدینگ، حذف شیب، انحنا و افزایش داده‌ها با شبکه‌های عصبی کانولوشنی دروازه‌دار و شبکه‌های عصبی بازگشتی ترکیب می‌کند. این طراحی یکپارچه به طور چشمگیری دقت و استحکام سیستم‌های تشخیص دست‌نویس را از طریق بهینه‌سازی کیفیت ورودی پیش از پردازش شبکه، افزایش می‌دهد. رویکرد جدید مزایای قابل توجهی از جمله بهبود دقت شناسایی و کاهش تأثیر نویز و الگوهای نامربوط را ارائه می‌دهد. نتایج آزمایشات انجام شده بر روی مجموعه داده IAM نشان می‌دهد که مدل پیشنهادی به نرخ

خطای کاراکتر (CER) ٪۴ دست یافته است، که نشان‌دهنده بهبود قابل توجه ٪۱۲ نسبت به روش‌های پیشرفته موجود می‌باشد.

مدلی که ارائه گردیده است قادر خواهد بود که بلوک‌های کلمات با اندازه‌های متفاوت همچون آدرس‌ها و اسامی افراد را تشخیص دهد و در مقایسه با روش‌های اخیر نتیجه‌ای بهتر به‌همراه دارد.

در ادامه‌ی این پژوهش، در فصل ۲ کارهای مربوطه ارائه می‌گردد. در فصل ۳ روش‌های تحقیق انجام‌شده شرح داده خواهد شد. در فصل سوم، در قسمت اول (بخش ۳-۱) ساختار کلی سیستم به‌تفصیل آورده شده است و در قسمت دوم و سوم فصل ۳ (بخش ۳-۲ و ۳-۳) روش‌های ارائه‌شده‌ی اول و دوم ارائه می‌گردد. در فصل چهارم نتایج، بحث و گفتگو را طرح گردیده است که نتایج پیاده‌سازی و جزئیات آن آمده است. در فصل پنجم نتیجه‌گیری و کارهای آینده آمده است؛ و درنهایت در فصل ۶ مراجع تحقیق، قرارگرفته‌اند.

۲- مروری بر کارهای گذشته

معمولاً برای اینکه شناسایی دست‌نوشته ساده‌تر شود، بر روی‌داده‌های دست‌نوشته نرمال‌سازی انجام می‌گیرد. آلمازان و همکاران [۲]، تصاویر کلمات مختلف را با یکدیگر مقایسه کرده و با نگاشت این تصاویر و برچسب‌های آن‌ها به فضایی جدا، عمل پیش‌بینی را انجام می‌دهند. آن‌ها نشان می‌دهند که در این فضای نگاشت‌شده، با جستجوی کلماتی که حاوی متنی مشابه هستند، قادر خواهند بود به‌راحتی کلمات مشابه، مستقل از نویسنده و نوع نوشتاری آن‌ها را پیش‌بینی کنند. دریو و همکاران [۲۸] همانند کوزلسکی[۲۹]، از مدل مخفی مارکوف استفاده کرده‌اند و نشان داده‌اند که مدل مخفی مارکوف گوسی از حالت معمولی آن کارایی بهتری دارد. کانگ و همکاران [۳]، محتوا و سبک ویژگی تصاویر ورودی را از یکدیگ جدا کردند. آن‌ها با استفاده از یک فرآیند مولد و تشخیص دهنده‌ی دست‌خط، ویژگی‌های نوشتاری را بدون وابستگی به نویسنده استخراج کردند به طوری که کارائی تشخیص دست‌خط در صورت وجود چندین نویسنده بهبود داده می‌شود. کوماری و همکاران [۴]، مکانیسم توجه[۲۰] را به شبکه‌های عصبی اضافه کردند، همینطور با استفاده از یک تابع فازی وزن‌های بالاتری به قسمت‌های مورد توجه متن درون تصاویر در نظر گرفتند و از این طریق به کارائی بهبود بخشیدند. موندال و همکاران [۶]، مدلی را بر اساس شبکه‌های عصبی عمیق ارائه داند که قابلیت تقسیم‌بندی و شناسائی کاراکتر در تصاویر کلمات دست‌نوشته را داراست. در این مدل با تعداد کم داده‌های آموزش، عملیات شناسائی حروف انجام پذیر می‌باشد. کاس و همکاران [۸]، علاوه بر استفاده از مکانیسم توجه، با استفاده از وزن‌های از پیش تعیین‌شده به دست‌آمده از تصاویر متنی مصنوعی، فرآیند آموزش را بهبود بخشیدند. بلوچ و همکاران [۷]، کار شناسایی حروف دست‌نویس را با مقایسه شبکه‌های عصبی بازگشتی و لایه‌های پرسپترون چندجهته[۲۱] (روش استاندارد برای تشخیص گفتار) انجام دادند و نشان می‌دهند خطای کلمه با استفاده از هر دو مدل قادر به بهبود می‌باشد. یکی از روش‌هایی که برای ساده‌سازی دست‌نوشته انجام می‌گیرد، رفع انحنا و رفع شیب حروف می‌باشد. به این صورت شکل حروف دست‌نوشته توسط نویسنده‌های مختلف تا حدودی شبیه یکدیگر شده و راحت‌تر توسط طبقه‌بند[۲۲] قابل‌شناسایی می‌باشد. در زمینه رفع شیب و انحنا حروف، وینکارلی و همکاران [۳۰]، الگوریتمی ارائه می‌دهند که ابتدا فرایند حذف شیب و سپس رفع انحنا را انجام می‌دهد. آن‌ها برای این کار ابتدا ناحیه اصلی تصویر را تخمین می‌زنند و با روش آستانه گذاری اوتسو[۲۳] شیب را تخمین می‌زنند. سپس با تشکیل یک هیستوگرام، انحنای حروف تخمین زده می‌شود. داتچ و همکاران [۹]، از شبکه‌های عصبی بازگشتی کوتاه درازمدت برای استخراج ویژگی‌های پیچیده در سطح خطوط استفاده کرده‌اند. کوزلسکی و همکاران [۳۱]، از رویکرد مدل‌های مخفی مارکوف با اعمال پیش‌پردازش‌های مختلف بهره برده‌اند. آقای آشورپور [۱۰] و همکاران، برای بازشناسی حروف مجزای فارسی از مدل مخفی مارکوف به جهت طبقه‌بندی و شناسایی حروف استفاده کرده‌اند. در این روش ابتدا پیش‌پردازش‌هایی بر روی‌داده‌ها انجام‌گرفته است، ویژگی‌ها استخراج‌شده است و سپس آموزش داده می‌شوند. اهلاوات [۳۲] و همکاران، با تغییر در پارامترهای شبکه‌های عصبی کانولوشنی به شناسایی ارقام دست‌نویس پرداختند. آن‌ها از بهینه‌ساز[۲۴] آدام[۲۵] برای طبقه‌بندی ارقام استفاده کردند و نشان دادند از سایر الگوریتم‌های بهینه‌ساز مبتنی بر گرادیان بهتر می‌باشد. بورا [۳۳] و همکاران، برای شناسایی حروف دست‌نوشته، شبکه‌های کانولوشنی را با طبقه‌بند رمزگذاری خروجی تصحیح خطا[۲۶] ترکیب کردند که شبکه عصبی کانولوشنی برای استخراج ویژگی و الگوریتم رمزگذاری خروجی تصحیح خطا به‌عنوان طبقه‌بند استفاده‌شده است. همچنین آن‌ها برای پیدا کردن شبکه عصبی کانولوشنی مناسب، سه شبکه معروف LeNET، AlexNet و ZfNet را مورد ارزیابی قراردادند که به این نتیجه رسیدند که LeNET نرخ دقت کمی را به دست می‌دهد، بنابراین با اضافه کردن لایه حذف تصادفی نورون[۲۷] و تابع واحد یک‌سوشده‌ی خطی[۲۸] بعد از لایه متراکم[۲۹] یا تماماً متصل[۳۰] اول، دقت را بهبود بخشیدند. همین‌طور نشان دادند که برای شناسایی حروف دست‌نوشته ترکیب شبکه AlexNet با طبقه‌بند رمزگذاری خروجی تصحیح خطا بهترین نتیجه را به دست

می‌دهد. پوزناسکی [۳۴]، از شبکه‌های کانولوشنی عمیق برای شناسایی کلمات با واژگان معین استفاده کرده است. اگرچه این رویکرد در روش‌هایی که واژگان ثابت و معین دارند نتیجه بهینه‌ای دارد، ولی با افزایش اندازه واژگان، کارایی را تحت تأثیر خود قرار می‌دهد. به‌ویژه، ممکن است اندازه دیکشنری برای اسامی خیابان‌ها، شماره تلفن‌ها و نام و نام خانوادگی افراد نامحدود باشد. در کار یوسف [۱۱]، محققان برای شناسایی دست‌نوشته از یک شبکه عصبی پیشخور[۳۱] استفاده کرده‌اند. همچنین آن‌ها به‌عنوان پیش‌پردازش از روش‌های افزونگی داده بهره برده‌اند. کاربون [۱۵]، با استفاده از شبکه‌های عصبی بازگشتی، یک سیستم تشخیص دست‌نوشته برخط طراحی کرده است و برای شناسایی زبان‌های مختلف از آن استفاده می‌کند. ایشان [۱۵]، با استفاده از منحنی بیز[۳۲] که در آن توالی ورودی کوتاه‌تر است به عمل تشخیص نهایی سرعت می‌بخشد. چن [۲۵]، از یک سیستم چندمنظوره به نام MuLTReNets استفاده کرده است و آن را به‌صورت یکجا برای تشخیص دست‌نوشته اعمال کرده است که شامل دو نسخه، یکی برای تشخیص حروف چند زبانی[۳۳] و دیگری برای تشخیص عبارات آبشاری[۳۴] استفاده‌شده است. از آنجایی که بلوک‌های کلمات دست‌نوشته ابعاد متفاوتی را دارا می‌باشند، می‌توان از روش‌هایی مثل شبکه‌های عصبی کانولوشنی کاملاً متصل بهره برد. کارتیک [۱۶]، نشان داده است که ترکیب شبکه‌های عصبی کانولوشنی کاملاً متصل با شبکه‌های عصبی بازگشتی نسبت به استفاده تنها یکی از آن‌ها کارایی بهتری را نتیجه می‌دهد. در این زمینه لانگ و چن [۱۷], [۱۸]، به‌خوبی نشان می‌دهند که این شبکه‌ها تصاویر با اندازه‌های گوناگون را به‌عنوان ورودی دریافت می‌کنند و قطعه‌بندی نواحی را به‌صورت هم‌زمان برای تشخیص و جداسازی انجام می‌دهند.

در تحقیق [۲۶], [۳۵], [۳۶] نویسندگان از شبکه‌های عصبی کانولوشنی دروازه‌دار برای شناسایی حروف تایپی و دست‌نویس انگلیسی و روسی بهره برده اند. دلیل استفاده آن‌ها از مکانیسم دروازه‌ای، استخراج ویژگی‌های مرتبط‌تر و پایین آوردن تعداد پارامترهای شبکه بوده است. توچا و همکاران [۲۷]، از شبکه‌های عصبی کاملاً کانولوشنی، به‌تنهایی در شناسایی هوشمند دست‌نوشته استفاده کرده‌اند. آن‌ها ابتدا با بهره‌گیری از یک شبکه کانولوشنی که شبیه به معماری VGG است، دنباله ورودی را نرمال کرده اند و سپس با استفاده از یک شبکه تماماً کانولوشنی شناسایی حروف را به انجام رسانده اند. شبکه آن‌ها قادر به شناسایی لغات و اسامی نام کوچک، نام خانوادگی و نام خیابان‌ها می‌باشد. آن‌ها برای بهبود خطای سطح حرف و کلمه، از روش یادگیری انتقالی[۳۵] استفاده کرده اند و در انتهای آموزش، عمل پس‌پردازش را با استفاده از یک برنامه‌نویسی پویا[۳۶] بر روی حروف شناسایی‌شده انجام داده اند.

در [۳۷] نویسنده برای تشخیص دست‌نویس در سطح کلمه از تعداد شبکه عصبی کانولوشنی و بازگشتی استفاده کرده است و تأکید بر استفاده از تعداد بیشتری از لایه‌های بازگشتی داشته است. در تحقیق [۳۸]، نویسندگان یک مدل چندرسانه‌ای پیشنهاد می‌دهند که دانش معنایی حاصل از مدل‌های متنی از پیش آموزش‌داده‌شده را به مدل وارد می‌کند و به بهینه‌سازی نمایش معنایی تصاویر کلمات دست‌نویس کمک می‌کند. در [۳۹] نویسندگان از مکانیزم توجه چند سری[۳۷] در مراحل پردازش، به تشخیص کاراکترها و یادگیری وابستگی‌های زبانی میان توالی کاراکترها می‌پردازند. این مدل محدود به واژگان از پیش تعریف‌شده نیست و قادر به شناسائی کلمات خارج از واژگان آموزشی نیز می‌باشد.

در مقاله [۲۸]، از مدل‌های ترکیبی شبکه عصبی چند لایه (MLP) و مدل پنهان مارکوف (HMM) استفاده می‌کنند تا با استفاده از ویژگی‌های غیرخطی و وابستگی‌های طولانی‌مدت، بهبودهای قابل‌توجهی در تشخیص دست‌نویس به‌دست آورند. رویکرد آن‌ها در بکارگیری آموزش MMI/MPE و استفاده از مدل‌های هیبریدی MLP/HMM برای تشخیص پیوسته دست‌نویس، نشان‌دهنده‌ی تلاش برای تلفیق روش‌های آماری با مدل‌های شبکه‌های عصبی جهت بهبود دقت است. در [۵] نیز از مدل‌های ترکیبی HMM/ANN برای تشخیص دست‌نویس استفاده می‌کنند، اما از تیکنیک‌هایی مانند حذف شیب و نرمال‌سازی اندازه متن نیز استفاده می‌شود. در [۲۹] نیز از یک سیستم مبتنی بر HMM استفاده شده است. این سیستم از نرمال‌سازی تصویر مبتنی بر لحظات، انطباق دستخط نویسنده و استخراج ویژگی‌های تفکیکی استفاده می‌کند و به طور ویژه برروی تشخیص واژگان در سطح کلمه تمرکز دارد. [۲] به مشکلات شناسائی کلمات و جستجوی آن‌ها در تصاویر می‌پردازد. نویسندگان با ایجاد فضایی مشترک میان تصاویر کلمات و رشته‌های متنی از طریق ترکیبی از تعبیه برچسب و یادگیری ویژگی‌ها، به روشی سریع و کم‌بعد برای تشخیص کلمات و جستجو دست می‌یابند. این رویکرد در مقایسه با روش‌های دیگر، به‌ویژه به‌دلیل کاهش پیچیدگی محاسباتی و امکان پردازش سریع، دارای مزایای قابل‌توجهی است. در [۹] نویسندگان بر روی استفاده از شبکه‌های عصبی بازگشتی با حافظه بلند کوتاه‌مدت متمرکز است و از رویکردهای جدیدی برای بهبود سرعت آموزش مدل‌ها از طریق پردازش‌های مینی‌بچ و تقسیم توالی‌ها بهره می‌گیرد. این مدل‌ها با استفاده از داده‌های عمومی و به‌کارگیری پردازش موازی، تشخیص دست‌نویس را انجام می‌دهد. در [۷]، آن‌ها به مقایسه مدل‌های شبکه‌های عصبی بازگشتی دوجهتی (Bidirectional LSTM) با

شبکه‌های عمیق چندلایه‌ای (DeepMLP) برای تشخیص دست‌نویس در واژگان‌های بزرگ پرداخته و نشان می‌دهد که هر دو مدل می‌توانند نتایجی قابل‌مقایسه با روش‌های مرسوم در این حوزه به‌دست آورند. این پژوهش بر تفاوت‌ها و شباهت‌های موجود بین این دو رویکرد و کاربرد آن‌ها در تشخیص دست‌نویس تأکید می‌کند. در پژوهش [۲۷] رویکردی مبتنی بر شبکه‌های تماماً کانولوشنی برای تشخیص دست‌نویس پیشنهاد می‌دهند که با حذف نیاز به هم‌ترازی نمادها، مشکلاتی مانند اصلاح هم‌ترازی را کاهش می‌دهد. این مدل با استفاده از معماری چندحالتی و بدون وابستگی به فرهنگ لغت از پیش تعریف‌شده، به نتایجی هم‌تراز با روش‌های مبتنی بر واژگان در تشخیص دست‌نویس می‌رسد. پژوهش [۳] با تمرکز بر عدم وابستگی به سبک نوشتاری نویسندگان، روشی را برای جداسازی ویژگی‌های محتوا و سبک از تصاویر کلمات دست‌نویس پیشنهاد می‌دهند. این روش با به‌کارگیری فرآیندهای تولیدی برای انتقال ویژگی‌های سبک و دستیابی به نمایش‌های مستقل از سبک نوشتاری، به بهبود دقت مدل‌های تشخیص دست‌نویس می‌پردازد. پژوهش [۴] به معرفی یک رویکرد مبتنی بر مدل‌های توالی به توالی می‌پردازند که از روش تقسیم‌بندی شبه فازی برای تشخیص کلمات استفاده می‌کند. این روش با بهبود تقسیم‌بندی و استفاده از مکانیزم توجه، به بهینه‌سازی شناسایی کاراکترها می‌پردازد. پژوهش [۶] از یک مدل مبتنی بر YOLOv3 استفاده می‌کنند که بدون نیاز به واژگان (lexicon-free) قادر به شناسایی متون دست‌نویس است. این مدل در مقایسه با اغلب روش‌ها، از واژگان‌های از پیش تعریف‌شده استفاده نمی‌کند و به همین دلیل قادر به تشخیص کلمات خارج از دایره‌ی واژگان آموزشی است. پژهش [۳۹] رویکردی غیر ترتیبی را برای شناسایی دست‌نویس ارائه کرده و با استفاده از لایه‌های توجه، در هر دو مرحله بصری و متنی، به‌طور کلمات خارج از واژگان را شناسایی می‌کند. بر خلاف روش‌های قبل پژوهش [۸] و [۳۷]، بر شناسایی واژگان از پیش تعیین‌شده متمرکز هستند و در رویکرد آن‌ها، استفاده از واژگان‌های از پیش تعریف‌شده به بهبود دقت کمک می‌کند. در این مدل‌ها، تمرکز بیشتر بر شناسایی دقیق کلمات متداول و کاهش خطای شناسایی در چارچوب واژگان‌های موجود است .

۳- معماری پیشنهادی

روشی که در این مقاله ارائه گردیده است، بر روی بلوک‌هایی از کلمات که از نویسه‌ها و نمادهای مختلف تشکیل‌شده‌اند تمرکز دارد. ورودی مدلی که ارائه شده است، تصویر می‌باشد. یک ورودی، از یک تصویر خاکستری‌رنگ شامل دنباله‌ای از علائم مختلف یک‌بعدی از نویسه‌ها و یا نمادها[۳۸] تشکیل‌شده است. لزوماً مدلی که ارائه شده است تنها به حروف لاتین محدود نمی‌شود و در صورت وجود مجموعه داده‌های مختلف، قابلیت تشخیص با میزان خطای بسیار کمی را داراست و ازنظر مقایسه با روش‌های اخیر نتیجه مطلوبی به دست می‌دهد.

۳-۱- پیش‌پردازش

روشی که در این پژوهش ارائه شده است، از دو بخش اصلی پیش‌پردازش و فرآیند آموزش[۳۹] (که خود از دو روش مجزای تماماً کانولوشنی و هیبریدی شامل لایه‌های کانولوشنی و بازگشتی استفاده کرده است) تشکیل‌شده است که در بخش‌های بعدی به‌تفصیل راجع به آن‌ها بحث خواهد شد. در این روش ابتدا داده[۴۰] که از نوع تصویر می‌باشد به مدل کلی وارد شده است. قبل از فرآیند آموزش، فرآیند پیش‌پردازش هست که در این قسمت تصویر با توجه به نسبت طول به عرض آن به یک اندازه متعارف تبدیل شده است و بر روی آن لایه‌گذاری صورت می‌پذیرد؛ و در ادامه شیب و انحنای حروف موجود در آن تشخیص داده‌شده و نرمال‌سازی می‌شوند. افزونگی داده انجام می‌گیرد. در قسمت فرآیند آموزش، دو رویکرد مجزا ارائه شده است که اولی از شبکه‌ای با لایه‌های تماماً کانولوشنی تشکیل‌شده است و دومی تحت عنوان مدل دروازه‌دار ترکیبی[۴۱] آمده است که ترکیبی از بلوک کانولوشنی و بازگشتی می‌باشد.

ابتدا در قسمت پیش‌پردازش به نرمال‌سازی داده‌های تصویری پرداخته شده است به صورتی که کاربردی فراتر از تصاویر جداشده‌ی ساده داشته باشد، و بتوان از این مدل در تصاویری که اعوجاج بسیار زیادی دارند نیز استفاده کرد. این کار ابتدا با استفاده از یک نرمال‌سازی با تغییر اندازه تصاویر شروع می‌شود. با رسیدن به اندازه مدنظر، با یک الگوریتم صاف‌سازی حروف انحنای اضافی نمادها رفع می‌شوند. این کار کمک می‌کند تا نمادهای دست‌نویس که توسط نویسنده‌های مختلف نوشته‌شده‌اند شبیه به یکدیگر شوند. برای این کار از روشی که در پژوهش [۳۰] معرفی‌شده است استفاده شده است. این الگوریتم ابتدا ورودی را با میزان مشخصی از مقدار پیکسل‌های تصاویر برش می‌دهد سپس با استفاده از یک تابع (که از مرزهای خطوطی که تقریباً عمودی می‌باشند به‌دست‌آمده است)، این برش‌ها را مورد ارزیابی قرار می‌دهد. این روش نیازی به هیچ تنظیم پارامتری ندارد و به‌صورت جداگانه می‌تواند در هر دو سطح کلمه و جمله استفاده گردد؛ و در انتهای فرآیند پیش‌پردازش، افزونگی داده انجام می‌گیرد.

ازاین‌رو، در قسمت فرآیند آموزش، از دو روش متفاوت شبکه عصبی تماماً کانولوشنی و شبکه هیبریدی استفاده شده است و نتایج بر

اساس دو پارامتر خطای بازشناسی حرف[۴۲] و خطای کلمه[۴۳] گزارش شده است.

برای این‌که نتایج مطلوبی از شبکه دریافت شود لازم هست فرآیند و یا فرآیندهایی بر روی‌داده‌های خام پایگاه داده انجام گیرد که در ادامه بررسی می‌شوند. باهدف نرمال‌سازی داده‌ها، پیش‌پردازش بر روی آن‌ها انجام می‌گیرد که توضیح روند آن در ۳-۱-۱ الی ۳-۱-۳ آمده است.

۳-۱-۱- تغییر اندازه بلوک کلمه با حفظ نسبت طول به عرض

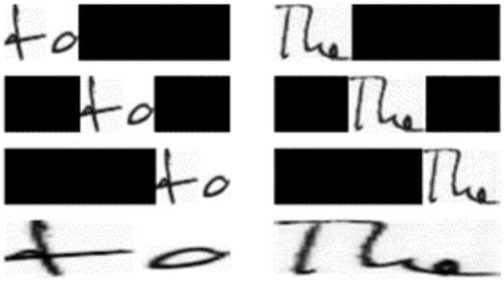
در این پژوهش براساس [۴۰] اندازه بلوک‌های تصویر کلمه ۱۲۸*۳۲ انتخاب می‌شود تا مشکلات کم بودن نمونه‌های دادگان و پیش‌بینی‌های آن پوشش داده می‌شود. این مورد را باید در نظر گرفت که علیرغم این‌که تصاویر مربعی هنگام تغییر اندازه دچار آن‌چنان اعوجاجی نمی‌شوند، این امر برای تصاویر مستطیل شکل صادق نیست. درهرصورت تغییر اندازه تصاویر به‌اندازه‌ای یکتا و ثابت یک نوع الزام برای فرآیند دسته کوچک[۴۴] می‌باشد، بنابراین می‌بایست این تغییر شکل تصاویر به‌گونه‌ای انجام گیرد که نه‌تنها محتویات تصویر مورد اعوجاج قرار نگیرد بلکه نسبت طول به عرض آن‌ها ثابت باقی بماند.



شکل ۱: مقایسه تصاویر در حالت برابری و عدم برابری طول به عرض

همان‌طور که در شکل ۱ مشهود است برای این‌که میزان طول و عرض تصویر دست‌خوش تغییر واقع نشود، بر روی تصاویر ردیف اول لایه‌گذاری صورت گرفته است، درصورتی‌که در تصاویر ردیف دوم تنها تغییر سایز انجام شده است. به همین دلیل اندازه‌ی آن‌ها کشیده شده است و شکل تصویر اولیه مورد اعوجاج قرارگرفته است.

برای اینکه نسبت طول به عرض تصاویر حفظ شود می‌توان از سه نوع لایه‌گذاری چپ، وسط و راست استفاده کرد. شکل ۲ نمای کلی از لایه‌گذاری‌های مختلف را نمایش می‌دهد.

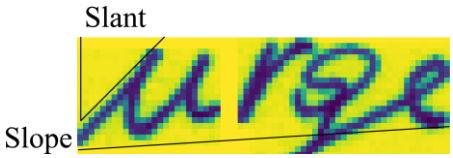


شکل ۲: لایه‌گذاری‌های مختلف برای نمونه ورودی

در ردیف اول دو نمونه از کلمات پایگاه داده IAM آمده است که لایه‌گذاری از چپ بر روی آن‌ها انجام‌گرفته است و به همین ترتیب در ردیف دوم و سوم لایه‌گذاری وسط و راست و در آخرین ردیف دو نمونه از همان کلمات با تغییر شکل و بدون در نظر گرفتن نسبت طول به عرض انجام‌شده است که ملاحظه می‌کنیم تصویر دچار دگرگونی نسبت به حالت اولیه شده است.

۳-۱-۲- رفع انحنا و شیب حروف

شیب عبارت است از زاویه بین جهت افقی و جهت خطی که حروف با آن تراز شده‌اند. همچنین، کجی یا انحنای حروف عبارت است از زاویه بین جهت عمودی و حروفی که در بهترین حالت می‌بایست عمودی باشند. در شکل ۳، شیب و انحنا مشهود می‌باشند.



شکل ۳: نمایش شیب و کجی

در این پژوهش، برای نرمال‌سازی حروف دست‌نوشته شکسته از الگوریتم وینسیارلی و همکاران[۳۰] استفاده شده است که روشی مناسب برای حذف شیب و انحنای زیاد حروف شکسته می‌باشد. با استفاده از این روش، درصد شناسایی علائم بهبود داده شده است.

۳-۱-۳- افزونگی داده

معمولاً هنگام آموزش یک شبکه عصبی برای مقاوم‌سازی شبکه نسبت به تغییرات دسته‌های کلاس‌ها و همین‌طور جلوگیری از بیش‌برازش از روش‌های دگرگونی یا تغییرات مصنوعی در داده‌ها استفاده می‌شود. مرسوم‌ترین روش‌ها شامل بریدن قسمتی از تصویر، انعکاس افقی تصویر، برگرداندن تصویر پیکسل‌ها و دگرگونی‌های نسبی مانند مقیاس‌گذاری و حرکت انتقالی است. در این تحقیق، از دو نوع افزونگی داده تغییرات نسبی[۴۵] و اعوجاج کشسان در حین زمان فرآیند آموزش و آزمون استفاده شده است. در این روش

حرکات دورانی تصاویر به ۵+ و ۵- درجه و حرکات برشی ۰.۵+ و ۰.۵- محدود شده است.

ازآنجایی‌که دستخط انسان به دلیل عدم وارد آمدن نیروی عضله یکسان بهنگام نوشتن باعث به‌وجودآمدن نوسان زیادی می‌باشد، با استفاده از اعوجاجات کشسان [۱۹] که برای افزونگی داده‌ها اعمال می‌گردد تا حدود خوبی می‌توان این تغییرات را پیدا کرد. ایده اصلی این روش این است که یک جابجایی تصادفی (که با فیلتر گوسی و انحراف از معیار سیگما و یک مقدار ثابت تراز شده است) را تولید کند. سپس محاسبه مکان جدید هر پیکسل از طریق درون‌یابی انجام می‌گیرد. در شکل ۴ نمایی از افزونگی داده اعمال‌شده را می‌بینیم.



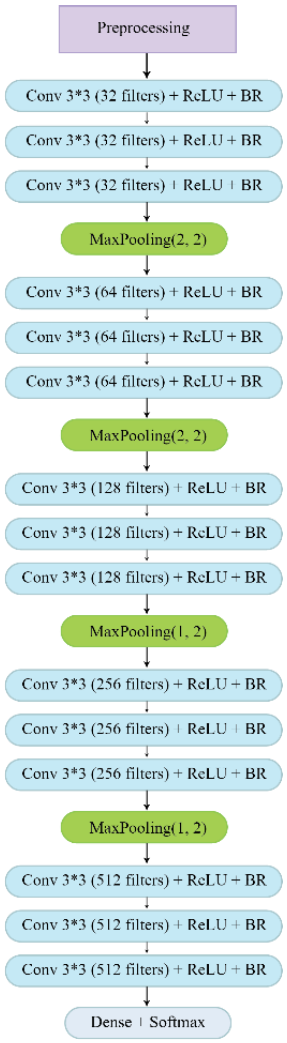
شکل ۴: اعمال افزونگی داده بر روی تصاویر

در شکل ۴ دو نمونه از تصاویر پایگاه داده IAM را با لایه‌گذاری سمت چپ نمایش می‌دهد. ردیف اول مربوط به تصاویر بدون اعوجاج را نشان می‌دهد، ردیف دوم همان تصاویر موجود هستند که بر روی آن‌ها الگوریتم رفع انحنای حروف اعمال‌شده است و در دو ردیف آخر فرآیند افزونگی داده با استفاده از تغییرات نسبی و اعوجاجات کشسان انجام گرفته‌است.

با استفاده از پیش‌پردازش پیشنهادی (فرآیندهای بخش ۳-۱-۱ الی ۳-۱-۳) اعوجاج کمتری در شبکه پیدا می‌شود، بنابراین در مدل‌های بخش بعد (بخش ۳-۲ و ۳-۳)، یک گام به‌عنوان پیش‌پردازش درنظر گرفته شده است.

۳-۲- روش پیشنهادی اول: شبکه‌های عصبی تماماً کانولوشنی FCOCR[46]

شکل ۵، اولین معماری پیشنهادی با نام FCOCR را نمایش می‌دهد. در این مدل ابتدا بر روی تصاویر، فاز پیش‌پردازش ارائه‌شده در بخش ۳-۱ اعمال می‌گردد. معماری که ارائه شده است شامل ۱۵ لایه کانولوشنی است که به جهت جلوگیری از ازدیاد بار محاسباتی آن از ۴ لایه پولینگ استفاده‌شده است. همچنین پس از هر لایه کانولوشنی از تابع فعال‌ساز رلو و نرمال‌سازی دسته‌ای باقابلیت نرمال‌سازی مجدد[47] استفاده شده است. در این معماری برای شناسایی حروف تنها از شبکه‌های عصبی کانولوشنی استفاده شده است و لایه‌های بازگشتی لحاظ نشده اند. این مدل قابلیت تشخیص بلوک کلمات و علائم با طول دلخواه با دقت مناسب را دارد.



شکل ۵: معماری ارائه‌شده روش اول: مدل متصل تماماً کانولوشنی FCOCR

نتایج پیاده‌سازی FCOCR برروی پایگاه داده IAM در جدول ۱ آمده است.

جدول ۱: نتایج شناسایی بلوک کلمات با معماری شکل ۵ و ۶

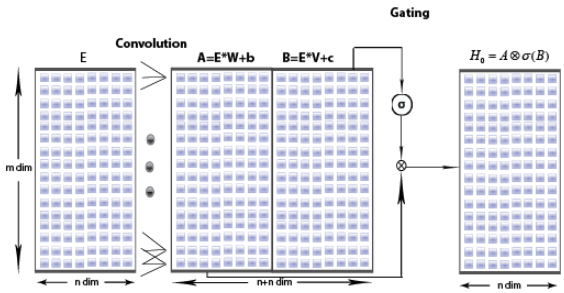
معماری	خطای کلمه (WER)	خطای حرف (CER)	زمان آموزش (ساعت)	تعداد پارامتر (میلیون)
مدل پیشنهادی اول (FCOCR)	۲۳/۰۸	۷/۲۰	۸	۷/۵

در ردیف اول جدول ۱ نتایج پیاده‌سازی اولین مدل پیشنهادی یا همان (FCOCR) (شکل ۵) گزارش‌شده است. با طراحی مناسب در

تعداد لایه‌ها و معماری شبکه، نرخ خطا در سطح کلمه ۲۳/۰۸ و کاراکتر ۷/۲۰ می‌باشد. تعداد پارامترهای این شبکه تنها ۷/۵ میلیون است که مقداری قابل قبول برای شبکه‌های شناسایی حروف محسوب می‌شود. برای بهبود خطای حروف و خطای کلمه در زیربخش (۳-۳) (روش پیشنهادی دوم) از شبکه های‌بازگشتی و شبکه‌های کانولوشنی دروازه‌دار استفاده می‌کنیم.

۳-۳- روش پیشنهادی دوم: شبکه‌های عصبی کانولوشنی دروازه‌ای (GCOCR[48])

در این بخش از پژوهش، یک معماری شبکه عصبی کانولوشنی بازگشتی ارائه شده است که از اهداف آن دقت بالا و کاهش هزینه محاسباتی می‌باشد. در این روش از شبکه عصبی کانولوشنی دروازه‌دار استفاده شده است که در آن تابع فعال‌ساز سیگموئید به نیمی از ویژگی‌ها اعمال می‌گردد و در ادامه با نیمی دیگر به‌صورت نقطه‌ای ضرب می‌شود (شکل ۶).

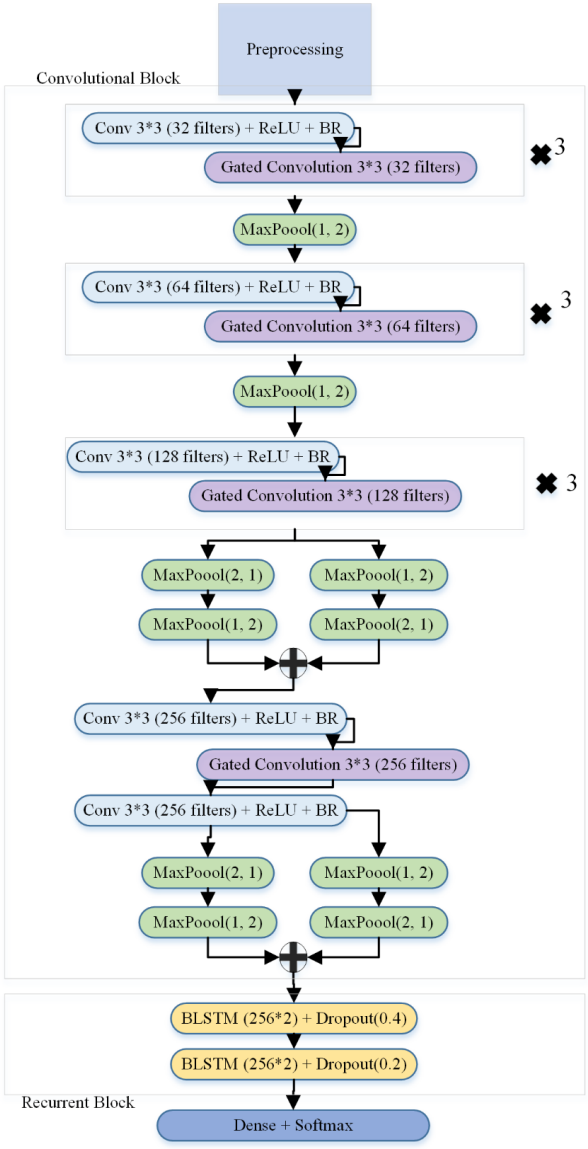


شکل ۶: ساختار لایه کانولوشنی دروازه‌دار

شکل ۶، ساختار لایه کانولوشنی دروازه‌دار را نمایش می‌دهد. پس از اعمال عملیات کانولوشن، خروجی بردار A توسط مقدار سیگموئید (σ) بردار B کنترل می‌شود. در شکل ۶ بردار B به‌عنوان دروازه خروجی استفاده‌شده است که در بردار A به‌صورت نقطه‌ای ضرب[49] می‌شود؛ بنابراین بردار E به بردار $H_0 = A \otimes \sigma(B)$ نگاشت می‌گردد.

از طریق این مکانیسم بار محاسباتی شبکه کمتر می‌شود و همچنان کارایی بهبود داده می‌شود (شکل ۷). این شبکه‌ها برای کنترل جریان اطلاعات، مکانیزم دروازه‌ای (مانند شبکه LSTM) در میان لایه‌های کانولوشنی را معرفی می‌کنند. این مکانیزم باعث می‌شود تا شبکه به اطلاعات مرتبط توجه کند و از پارامترها موثرتر استفاده کند.

در شبکه‌های کانولوشنی معمولی جریان اطلاعات به صورت ترتیبی از طریق لایه‌ها جریان می‌یابد و کنترلی بر روی حفظ و یا ازدست‌دادن اطلاعات وجود ندارد در صورتی که شبکه‌های دروازه‌دار برای مدلسازی دنباله‌ای مناسب هستند و می‌توانند وابستگی‌های بلندمدت و الگوهای متوالی در داده‌ها را به خوبی ضبط کنند.



شکل ۷: معماری روش پیشنهادی دوم: مدل GCOCR

به‌طورکلی، فرآیند دروازه‌ای برای کارایی مناسب شبکه‌های عصبی کانولوشنی بسیار ضروری می‌باشد. این واحدهای خطی دروازه‌ای مشکل ناپدیدشدن گرادیان‌ها را برای معماری‌های عمیق، با فراهم آوردن یک مسیر خطی و حفظ کردن قابلیت‌های غیرخطی گرادیان‌ها رفع می‌کند.

پس از نرمال‌سازی داده‌ها (اجرای روند پیش‌پردازش)، شبکه عصبی موردنظر تشکیل می‌گردد و آماده‌سازی داده‌ها برای ورود به مدل انجام می‌گیرد.

شکل ۷ معماری شبکه ارائه‌شده (مدل GCOCR) را نمایش می‌دهد (شبه کد معماری در بخش پیوست آمده است). در این شکل بلوک‌های کانولوشنی و بازگشتی به‌صورت پشته‌ای بر روی یکدیگر قرارگرفته‌اند. پس از هر لایه کانولوشنی یک بلوک کوچک با مکانیسم دروازه‌ای قرار داده‌شده است. تصویر ورودی به این شبکه ابتدا فاز پیش‌پردازش (توضیح داده‌شده در بخش ۳-۱) را می‌گذراند تا برای فرآیند لایه‌های شبکه آماده گردد.

این معماری، تحت عنوان مدل GCOCR از ۱۱ لایه کانولوشنی تشکیل‌شده است که در میان هرکدام از آن‌ها یک بلوک کوچک[۵۰] موجود است. این بلوک‌های کوچک از مکانیسم دروازه‌ای استفاده می‌کنند. در این مدل نگاشت‌های ویژگی[۵۱] دارای کانال ۳۲ برای ۳ بلوک کوچک اول، ۶۴ برای ۳ بلوک کوچک دوم، ۱۲۸ برای ۳ بلوک کوچک سوم و ۲۵۶ برای دو بلوک دروازه‌ای و کانولوشنی آخر می‌باشند. کرنل اعمال‌شده ۳*۳ در نظر گرفته شده است و همین‌طور، به‌تمامی لایه‌های کانولوشنی نرمال‌سازی مجدد دسته‌ای [۴۱] اعمال‌شده است. لایه حذف تصادفی (در هر انتشار نیمی از ورودی فیلترها صفر قرار می‌گیرد) پس از لایه‌های بازگشتی اعمال‌شده‌است. از لایه مقدار دهنده‌ی توزیع یکنواخت[۵۲] برای کرنل‌ها، به‌عنوان مقداردهی اولیه وزن‌ها و رلو به‌عنوان تابع فعال‌ساز استفاده شده است. خروجی شبکه دارای ۸۱ کلاس شامل حروف بزرگ و کوچک انگلیسی، اعداد و برخی نمادها مانند" ' ! # $ ٪ &) (*+-/ : ; ، ? به‌علاوه یک نماد خالی[۵۳] می‌باشد. همچنین به جهت حفظ هرچه بیشتر نگاشت‌های ویژگی از لایه‌های MaxPool با ابعاد ۱*۲ و ۲*۱ استفاده‌شده است. الحاق بانماد ⊕ لایه‌های MaxPool(1, 2) و MaxPool(2, 1) بعد از لایه کانولوشن دروازه‌ای با ۱۲۸ کانال و یکبار بعد از لایه کانولوشن با ۲۵۶ کانال انجام می‌گردد. استفاده از MaxPool(1, 2) و MaxPool(2, 1) به معنای اعمال لایه ادغام با گام‌های مختلف در ابعاد مختلف است، که می‌تواند ویژگی‌های مختلفی را دریافت نماید. ادغام نتایج این دو عملیات به مدل اجازه می‌دهد که اطلاعاتی را از جهت‌های افقی و عمودی به‌طور جداگانه جمع‌آوری کند.

از سوی دیگر MaxPool(2, 2)، ادغام با گام یکسان در هر دو بعد انجام می‌دهد که ممکن است منجر به از دست رفتن اطلاعات به دلیل کاهش فضائی بیشتر شود.

در مدل ما، استفاده از MaxPool(1, 2) و MaxPool(2, 1) و سپس ادغام نتایج با یکدیگر به جای استفاده از MaxPool(2, 2)، منجر به کاهش بیشتر خطا شده است.

بلوک بازگشتی این ماژول، شامل دولایه حافظه طولانی کوتاه‌مدت دوجهته به همراه لایه حذف تصادفی با نرخ ۰.۴ بعد از لایه بازگشتی اول و نرخ ۰.۲ بعد از لایه بازگشتی دوم می‌باشد. در ادامه با استفاده از لایه‌های کاملاً متصل و Softmax به شبکه اجازه می‌دهیم تا ویژگی‌های پیچیده‌تری را از داده‌ها را استخراج کرده و بهبود دقت و عمکرد شبکه را فراهم کنیم. توضیحات بیشتر در این خصوص در پیوست ب جدول ۴ تشریح گردیده است.

در جدول شماره ۲ نتایج پیاده‌سازی بر روی پایگاه داده IAM با معماری شرح داده‌شده آورده شده است.

جدول ۲: نتایج مدل‌های پیاده‌سازی‌شده بر روی پایگاه داده IAM.

معماری	خطای کلمه (WER)	خطای حرف (CER)	زمان آموزش (ساعت)	تعداد پارامتر (میلیون)
مدل پیشنهادی اول FCOCR	۲۳/۰۸	۷/۲۰	۸	۷.۵
مدل پیشنهادی دوم GCOCR	۸/۰۳	۴/۱۴	۱۱	۶

در جدول شماره ۲ نتیجه روش پیشنهادی اول: معماری تماماً کانولوشنی (شکل ۵) و نتیجه روش پیشنهادی دوم: (مدل ترکیبی دروازه‌دار-شکل ۷) آمده است. ملاحظه می‌کنیم که با ترکیب شبکه‌های عصبی کانولووشنی و استفاده از دولایه بازگشتی دو جهتی خطای سطح کلمه و حروف کاهش و بهبود داده شده‌اند. در این مدل از هیچ‌گونه مدل از پیش آموزش‌دیده‌ای استفاده نکرده‌ایم. در این مدل از شبکه‌های دروازه‌ای استفاده‌شده است که به خاطر ماهیت این نوع از شبکه‌ها دقت نیز افزایش‌یافته است. همان‌طور که در بخش‌های پیش اشاره شد این شبکه‌ها ویژگی‌های فضای شبکه در یک مکان را موردبررسی قرار داده و با مقایسه آن با فضای شبکه مجاورش تصمیم به حفظ و یا حذف آن ویژگی‌ها می‌کند.

۴- پیاده‌سازی و تحلیل نتایج

برای اندازه‌گیری کارایی روش‌های ارائه‌شده از دو معیار نرخ خطای حرف و نرخ خطای کلمه استفاده شده است. فرمول استفاده‌شده برای نرخ خطای حرف به‌صورت زیر تعریف می‌شود[۴۲]:

$$\text{نرخ خطای حروف}^{54} = \frac{S_c + D_c + I_c}{S_c + D_c + C_c} \quad (۱)$$

$$\text{نرخ خطای کلمات}^{55} = \frac{S_w + D_w + I_w}{S_w + D_w + C_w} \quad (۲)$$

در فرمول فوق S تعداد حروف جایگزین‌شده، D تعداد حروف حذف‌شده، I تعداد حروف درج‌شده، C تعداد حروف صحیح می‌باشد

و N تعداد کل حروف می‌باشد. درواقع دو مسئله یکی می‌باشند ولی پارامترهای ذکرشده یک‌بار در سطح حروف (اندیس c) و یک‌بار در سطح کلمه (اندیس w) مورداستفاده قرار می‌گیرند.

۴-۱- پایگاه داده مورد استفاده

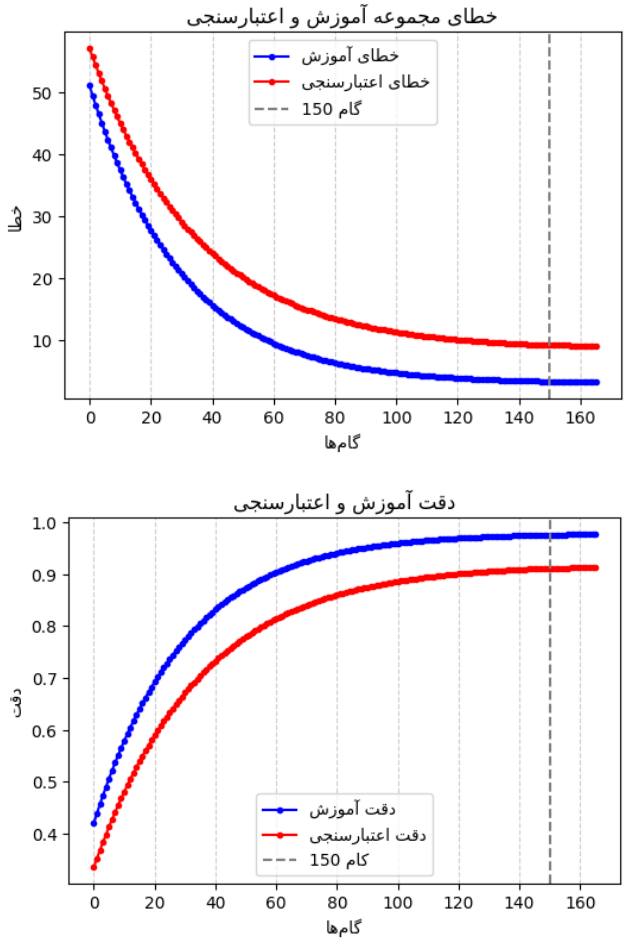
نتایج این تحقیق بر روی مجموعه داده حروف دست‌نویس IAM تست شده‌است. این پایگاه داده از ۶۵۷ نویسنده شامل ۱۱۵،۳۲۰ کلمه انگلیسی که اکثراً سرهم می‌باشند تشکیل شده است. این مجموعه داده شامل داده‌های آموزش، صحت و تست می‌باشد به طوری که افرادی که دست‌خط آن‌ها به عنوان نویسنده در قسمت آموزش استفاده شده‌اند، در سایر قسمت‌های صحت و تست آورده نشده اند. ما ۱۰۵،۶۷۴ داده را برای آموزش و ۴۸۲۳ داده برای صحت و ۴۸۲۳ داده برای مجموعه داده تست درنظر گرفتیم.

۴-۲- جزئیات شبیه‌سازی

برای شبیه‌سازی این پژوهش از کارت گرافیکی GeForce GTX 1060 و فریم‌وورک تنسورفلو[۵۶] برای کلیه آزمایش‌ها استفاده شده است. برای کار با برچسب‌ها نیز تعدادی لایه سفارشی‌سازی شده تعبیه گردیده است.

آموزش مدل با استفاده از الگوریتم بهینه‌ساز آدام با نرخ یادگیری[۵۷] ۰.۰۰۱ استفاده شده است. همچنین از نرمال‌سازی L2-norm با λ=0.0025 استفاده شده‌است. هدف از آموزش مدل به حداقل رسانی خطای صحت تاحد ممکن بوده است. برای جلوگیری از بیش‌برازش و بهبود عملکرد مدل، در صورتی که بهبودی در معیار صحت مشاهده نشود (بعد از ۸۰ گام)، از مکانیسم‌هایی مانند کاهش خطای یادگیری[۵۸] (با نرخ ۰.۲) و توقف زودهنگام[۵۹] (با patience=20) برای شرایط خاتمه آموزش استفاده شده‌است. مدت زمان آموزش روش پیشنهادی اول ۸ ساعت و روش پیشنهادی دوم ۱۱ ساعت بوده است.

نمودراهای همگرایی سیستم در شکل ۸ ارائه گردیده است، که نمودارهای کاهش خطا برای مجموعه آموزش و اعتبارسنجی نشان‌دهنده عملکرد مدل یادگیری عمیق طی ۱۶۵ گام هستند. در ابتدای آموزش (گام ۰)، خطای آموزش از مقدار ۵۱ و خطای اعتبارسنجی از مقدار ۵۷ شروع می‌شوند، که نشان‌دهنده دشواری اولیه مدل در یادگیری داده‌ها است. با پیشرفت آموزش، خطاها روند نزولی نمایی دارند، به طوری که خطای آموزش در تمام گام‌ها کمتر از خطای اعتبارسنجی باقی می‌ماند. این فاصله نشان‌دهنده یادگیری مؤثر مدل و جلوگیری از بیش‌برازش است. در گام ۱۵۰، خطای آموزش به مقدار ۳ و خطای اعتبارسنجی به مقدار ۸/۵ می‌رسند و پس از آن بهبود کمی مشاهده می‌شود که نشان‌دهنده نزدیک شدن مدل به همگرایی است. در جدول شماره ۳ نیز مقایسه نتایج به‌دست‌آمده مدل ارائه‌شده با سایر نتایج کارهای پیشین آورده شده است.



شکل ۸: نمودارهای کاهش Loss و افزایش دقت نسبت به افزایش epoch

در جدول شماره ۳، نتایج ارائه‌شده و کارهای پیشین در این حوزه آمده است. همان‌طور که مشهود می‌باشد، نتایج به‌دست‌آمده قابلیت مقایسه با سایر روش‌های پیشین را داراست. همچنین ستونی برای تعداد پارامترها در نظر گرفته شده است که نشان‌دهنده‌ی متغیرهای کارهای مشابه می‌باشد. روش ارائه شده‌ی دوم، نسبت به روش اول دارای دقت بالاتر بوده است. همچنین تعداد پارامترهای کمتری را داراست. در رویکرد دوم، از روش‌های پس پردازشی مانند تطبیق پیش‌بینی‌های شبکه با دیکشنری کلمات و یا استفاده از n-gram ها خودداری شده است ولی همچنان نتیجه قابل‌مقایسه با سایر کارهای این حوزه می‌باشد. تعداد پارامترهای مدل نیز تا حد ممکن بهینه می‌باشد به‌طوری که قادر بوده است نتیجه‌ای بهتر نسبت به کارهای

اخیر همین زمینه به دست آورد. همچنین در هیچ‌کدام از رویکردهای پیشنهادی روش اول و دوم از یادگیری انتقالی و تنظیم سازی استفاده‌نشده و مدل از پایه آموزش دیده است. چراکه تمرکز اصلی ما بر روی قدرت کارایی شبکه به‌تنهایی بوده است و تا جای ممکن از روش‌های جانبی بهبود کارآمدی شبکه خودداری شده است.

تحلیل مقایسه‌ای این تحقیق حاکی از آن است که روش پیشنهادی ارائه شده نسبت به مدل‌های موجود روی پایگاه داده IAM در حوزه تشخیص حروف از نظر هر دو نرخ خطای حرف و خطای کلمه عملکرد بهتری دارد. به طور خاص، در مقایسه با پژوهش [۸] که یکی از جدیدترین و برجسته‌ترین مطالعات مشابه است، میزان خطای حرف در GCOCR به میزان ۳/۷٪و میزان خطای کلمه به میزان ۳۷٪ کمتر است. همچنین، مطالعه [۲۷] که به عنوان یکی از قوی‌ترین تحقیقات در این حوزه شناخته می‌شود، نشان می‌دهد که خطای حرف GCOCR به میزان ۶/۸٪ و خطای کلمه آن به میزان ۴٪/۲ کمتر از این تحقیق است. از نظر تعداد پارامترها، GCOCR نسبت به تمامی تحقیقات مشابه، به‌جز [۲۹]، پارامترهای کمتری دارد با این حال، پژوهش [۲۹] در مقایسه با GCOCR دارای خطای بیشتری در هر دو سطح کاراکتر و کلمه است.

مدل پیشنهادی GCOCR با استفاده از ترکیبی از شبکه‌های عصبی کانولوشنی دروازه‌دار و بازگشتی، همراه با تکنیک‌های پیش‌پردازش، توانسته است مزایای قابل توجهی را نسبت به سایر پژوهش‌ها به دست آورد. پژوهش‌هایی مانند .Mondal et al[۶] و Kang et [۳۹] al از مدل‌های بسیار پیچیده‌ای مانند YOLOv3 و ترانسفورمرها استفاده کرده‌اند که به تعداد زیادی پارامتر نیاز دارند و در نتیجه نیاز به منابع محاسباتی بیشتری دارند. در مقابل، مدل GCOCR با استفاده از تنها ۶ میلیون پارامتر، به دقت بهتر از این مدل‌ها دست یافته است. بهینه‌سازی تعداد پارامترها در این مدل، نه تنها باعث کاهش زمان آموزش شده، بلکه بهره‌وری بیشتری از منابع محاسباتی به همراه داشته است. در مقایسه با مدل‌های سنگین‌تر مانند ترانسفورمرها، GCOCR برای کاربردهای واقعی و محیط‌هایی با منابع محدود محاسباتی مناسب‌تر است. همینطور در حالی که مقالاتی مانند .Kass et al[۸] و Nisa H [۳۷] از معماری‌های بازگشتی (LSTM) و کانولوشن‌های استاندارد بهره می‌برند، روش پیشنهاد شده با ترکیب شبکه‌های کانولوشنی دروازه‌دار و لایه‌های بازگشتی، امکان کنترل دقیق جریان اطلاعات در شبکه را فراهم کرده است. این مکانیسم به شبکه اجازه می‌دهد تا اطلاعات مهم و مرتبط را حفظ و اطلاعات غیرضروری را حذف کند، در حالی که مدل‌های بازگشتی استاندارد قادر به این سطح از کنترل اطلاعات

جدول ۳: مقایسه نتایج با کارهای مربوطه

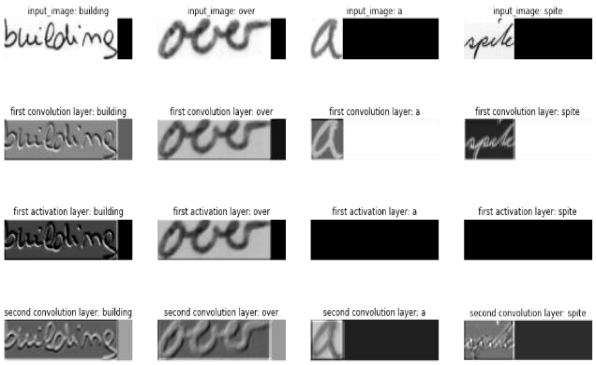
Model	WER	CER	#params	year
Dreuw et al. [۲۸]	۱۸/۸	۱۰/۱	-	۲۰۱۱
Boquera et al. [۵]	۱۵/۵	۶/۹۰	-	۲۰۱۱
Kozielski et al. [۲۹]	۱۳/۳	۵/۱	۰/۸ میلیون	۲۰۱۳
Almazan et al. [۲]	۲۰/۱	۱۱/۲۷	-	۲۰۱۴
Doetsch et al.[۹]	۱۲/۲	۴/۷	۱۰/۷ میلیون	۲۰۱۴
Bluche et al.[۷]	۱۱/۹	۴/۹		۲۰۱۴
Prtucha et al.[۲۷]	۸/۲۲	۴/۷	۱۹میلیون *	۲۰۱۸
Kang L et al. [۳]	۱۶/۳۹	۶/۴۳	-	۲۰۲۰
Bhattacharya et al. [۴]	۱۸/۶	۹/۷۰	-	۲۰۲۱
Mondal et al. [۶]	۲۹/۲۱	۹/۵۳	-	۲۰۲۲
Kass et al. [۸]	۱۲/۸۲	۴/۳۰	-	۲۰۲۲
Kang et al [۳۹]	۱۲/۹	۵/۱۸	۳۷ میلیون*	۲۰۲۲
Nisa H [۳۷]	-	۵/۱۶	۷ میلیون	۲۰۲۳
mann et al [۳۸].	۱۴/۳	۵/۵	-	۲۰۲۳
Proposed Method (GCOCR)	**۰۸/۰۳**	**۴/۱۴**	۶ میلیون	
*: تعداد پارامترهای محاسبه شده توسط ما				

نیستند. نتیجه این بهبود در معماری، کاهش نرخ خطای کاراکتر و کلمه به‌طور قابل‌توجهی است. همین‌طور اغلب پژوهش‌ها رویکردهای بدون واژگان (lexicon-free) مانند پژوهش‌های .Almazan et al[۲] و .Mondal et al[۶]، را برای تشخیص کلمات جدید و خارج از واژگان استفاده کرده‌اند. این پژوهش از تکنیک‌های پیش‌پردازش پیشرفته مانند رفع شیب، انحنا و افزونگی داده‌ها استفاده کرده‌است و توانسته است به طور مؤثرتری با مشکلات مربوط به تنوع نویسندگی و تغییرات سبک نوشتاری مواجه شود. این پیش‌پردازش‌ها باعث شده‌اند که شبکه بتواند نویسندگی‌های مختلف را با دقت بالاتری شناسایی کند، در حالی که بسیاری از مدل‌های دیگر به دلیل نبود چنین پیش‌پردازش‌هایی دچار افت دقت در مواجهه با تنوع نویسندگی هستند. در پژوهش Dreuw et .al[۲۸] و Boquera et al. [۵] و .Kozielski et al[۲۹] به دلیل استفاده از مدل‌های آماری مانند HMM، به نسبت ساده‌تر هستند و قادر به مدلسازی وابستگی‌های بلندمدت نیستند. مدل GCOCR با استفاده از شبکه‌های کانولوشنی دروازه‌دار، به‌طور موثرتری

ویژگی‌های پیچیده را استخراج می‌کند و با استفاده از لایه‌های بازگشتی، وابستگی‌های بلندمدت را نیز به خوبی مدل‌سازی می‌کند. این امر باعث می‌شود دقت و قابلیت یادگیری مدل در مواجهه با داده‌های پیچیده و نویسندگی‌های متنوع بهبود یابد.

۴-۳- بحث

به جهت ادراک بر روی کارایی فیلترهای یادگیرنده‌ی مدل ارائه‌شده، مطالعه کندوکاوی صورت گرفته و نتایج آن در شکل ۹ مشهود هست.



شکل ۹: بصری‌سازی و مطالعه کندوکاوی بر روی شبکه

شکل ۹ نشان‌دهنده‌ی اعمال نگاشت‌های ویژگی در لایه‌های مختلف می‌باشد. همچنین با تحقیق در این حوزه، دریافت شده است که بر روی تصاویر با رزولوشن بالاتر و تغییر ارتفاع پنجره از ۳۲ به ۶۴، درنتیجه نهایی تغییر چندانی اتفاق نخواهد افتاد.

در ردیف اول شکل ۹، چهار نمونه از تصویر ورودی آورده شده است. ردیف دوم و چهارم حاصل اعمال اولین و دومین لایه‌های کانولوشنی می‌باشد. ردیف سوم حاصل اعمال تابع فعال‌ساز رلو می‌باشد. لازم به ذکر است که مدل آموزش‌دیده توانسته است عملکرد خوبی را از خود نشان دهد. همچنین، ملاحظه می‌شود که برای مسأله شناسائی کاراکتر، پیش‌پردازش بلوک‌های حروف به همراه یک معماری مناسب در کنار هم باعث عملکرد بهتر درنتیجه نهائی خواهد شد.

۵- نتیجه‌گیری

در این تحقیق، یک الگوریتم شناسایی دستخط معرفی شده است که با استفاده از شبکه‌های کانولوشنی استاندارد، توالی داده‌های ورودی؛ مورد ارزیابی قرار داده شده است. این شناسایی هیچ‌گونه محدودیت واژگانی نداشته و می‌تواند شامل مجموعه علائم با طول بسیار بزرگ نیز باشد. این روش از راه‌های مؤثر با استفاده از نرمال‌سازی داده‌های ورودی و افزونگی داده برای آموزش شبکه استفاده می‌کند.

این روش قادر به شناسایی کلمات عامیانه مانند اسامی طولانی، نام افراد، شماره‌های تلفن و مخفف‌ها و سایر علائم و اختصارات است.

کارهای آتی در این حوزه می‌تواند شامل آزمایش شبکه با فیلترهای بزرگ‌تر و معرفی پردازش سلسله مراتبی برای به دست آوردن دانش در سطح جملات، پاراگراف و صفحات، استفاده از لایه‌های بازگشتی بیشتر با ساختارهای متفاوت، بکار بردن این پیکربندی برای سایر زبان‌هایی که منابع کمی به آن‌ها متصور است و سایر کاربردهایی که الگوریتم‌های بازگشتی در آن دارند، باشد. در آینده برای اصلاح اشتباهات هجی کلمات پس از فرآیند پیش‌بینی، حروف و یا علائم شناسایی‌شده توسط طبقه‌بند، می‌توانند مجدد پردازش شوند.

مراجع

[1] "Cross lingual handwritten character recognition using long short term memory network with aid of elephant herding optimization algorithm," *Pattern Recognit Lett*, vol. 159, pp. 16–22, Jul. 2022, doi: 10.1016/J.PATREC.2022.04.038.

[2] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Word spotting and recognition with embedded attributes," *IEEE Trans Pattern Anal Mach Intell*, vol. 36, no. 12, pp. 2552–2566, Dec. 2014, doi: 10.1109/TPAMI.2014.2339814.

[3] L. Kang, P. Riba, M. Rusiñol, A. Fornés, and M. Villegas, "Distilling Content from Style for Handwritten Word Recognition." 17th International Conference on Frontiers 2020.

[4] R. Bhattacharya, S. Malakar, F. Schwenker, and R. Sarkar, "Fuzzy-based pseudo segmentation approach for handwritten word recognition using a sequence to sequence model with attention," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12662 LNCS, pp. 582–596, 2021, doi: 10.1007/978-3-030-68790-8_45.

[5] S. España-Boquera, M. J. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez, "Improving offline handwritten text recognition with hybrid HMM/ANN models," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, no. 4, pp. 767–779, 2011, doi: 10.1109/TPAMI.2010.141.

[6] R. Mondal, S. Malakar, E. H. Barney Smith, and R. Sarkar, "Handwritten English word recognition using a deep learning based object detection architecture," *Multimed Tools Appl*, vol. 81, no. 1, pp. 975–1000, Jan. 2022, doi: 10.1007/s11042-021-11425-7.

[7] T. Bluche, H. Ney, and C. Kermorvant, "A comparison of sequence-trained deep neural networks and recurrent neural networks optical modeling for handwriting recognition," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8791, pp. 199–219, 2014, doi: 10.1007/978-3-319-11397-5_15/COVER.

[8] D. Kass and E. Vats, "AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.09390

[9] P. Doetsch, M. Kozielski, and H. Ney, "Fast and Robust Training of Recurrent Neural Networks for Offline Handwriting Recognition," *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, vol. 2014-December, pp. 279–284, Dec. 2014, doi: 10.1109/ICFHR.2014.54.

[۱۰] عاشورپور، محمد و زیارت‌بان، مجید. "بازشناسی برخط حروف مجزای دست‌نویس فارسی با استفاده از مدل مخفی مارکوف گسسته و ویژگی‌های مبتنی بر زبان فارسی،" *مجله علمی رایانش نرم و فناوری اطلاعات*، شماره ۶ (۲)، صفحات ۵۱-۶۸، پائیز ۱۳۹۶.

[11] M. Yousef, K. F. Hussain, and U. S. Mohammed, "Accurate, data-efficient, unconstrained text recognition with convolutional

neural networks," *Pattern Recognit*, vol. 108, Dec. 2020, doi: 10.1016/j.patcog.2020.107482.
[12] A. A. A. Ali and S. Mallaiah, "Intelligent handwritten recognition using hybrid CNN architectures based-SVM classifier with dropout," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3294–3300, Jun. 2022, doi: 10.1016/J.JKSUCI.2021.01.012.
[13] Z. Ullah and M. Jamjoom, "An intelligent approach for Arabic handwritten letter recognition using convolutional neural network," *PeerJ Comput Sci*, vol. 8, p. e995, May 2022, doi: 10.7717/PEERJ-CS.995/SUPP-2.
[14] P. Lopez-Rodriguez, J. G. Avina-Cervantes, J. L. Contreras-Hernandez, R. Correa, and J. Ruiz-Pinales, "Handwriting Recognition Based on 3D Accelerometer Data by Deep Learning," *Applied Sciences 2022, Vol. 12, Page 6707*, vol. 12, no. 13, p. 6707, Jul. 2022, doi: 10.3390/APP12136707.
[15] V. Carbune *et al.*, "Fast multi-language LSTM-based online handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 23, no. 2, pp. 89–102, Jun. 2020, doi: 10.1007/S10032-020-00350-4/TABLES/9.
[16] K. Dutta, P. Krishnan, M. Mathew, and C. V. Jawahar, "Improving CNN-RNN hybrid networks for handwriting recognition," *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, vol. 2018-August, pp. 80–85, Dec. 2018, doi: 10.1109/ICFHR-2018.2018.00023.
[17] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," 2015.
[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," pp. 1–12, Dec. 2014, doi: 10.48550/arxiv.1412.7062.
[19] P. Y. Simard, J. C. Platt, and D. Steinkraus, "Best Practices for Convolutional Neural Networks Streaming data View project Machine Teaching View project Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis," 2003, Accessed: Sep. 16, 2022. [Online]. Available: https://www.researchgate.net/publication/2880624
[20] M. Jaderberg, K. Simonyan, A. Zisserman, and koray kavukcuoglu, "Spatial Transformer Networks," *Adv Neural Inf Process Syst*, vol. 28, 2015.
[21] D. X. Zhou, "Universality of deep convolutional neural networks," *Appl Comput Harmon Anal*, vol. 48, no. 2, pp. 787–794, Mar. 2020, doi: 10.1016/J.ACHA.2019.06.004.
[22] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language Modeling with Gated Convolutional Networks," Jul. 17, 2017, *PMLR*. Accessed: Sep. 16, 2022. [Online]. Available: https://proceedings.mlr.press/v70/dauphin17a.html
[23] Y. N. Dauphin and D. Grangier, "Predicting distributions with Linearizing Belief Networks," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, Nov. 2015, doi: 10.48550/arxiv.1511.05622.
[24] M. M. Misgar, F. Mushtaq, S. S. Khurana, and M. Kumar, "Recognition of offline handwritten Urdu characters using RNN and LSTM models," *Multimedia Tools and Applications 2022*, pp. 1–24, Jun. 2022, doi: 10.1007/S11042-022-13320-1.
[25] Z. Chen, F. Yin, X. Y. Zhang, Q. Yang, and C. L. Liu, "MuLTReNets: Multilingual text recognition networks for simultaneous script identification and handwriting recognition," *Pattern Recognit*, vol. 108, p. 107555, Dec. 2020, doi: 10.1016/J.PATCOG.2020.107555.
[26] A. Abdallah, M. Hamada, and D. Nurseitov, "Attention-Based Fully Gated CNN-BGRU for Russian Handwritten Text," *Journal of Imaging 2020, Vol. 6, Page 141*, vol. 6, no. 12, p. 141, Dec. 2020, doi: 10.3390/JIMAGING6120141.
[27] R. Ptucha, F. Petroski Such, S. Pillai, F. Brockler, V. Singh, and P. Hutkowski, "Intelligent character recognition using fully convolutional neural networks," *Pattern Recognit*, vol. 88, pp. 604–613, Apr. 2019, doi: 10.1016/j.patcog.2018.12.017.
[28] P. Dreuw, P. Doetsch, C. Plahl, and H. Ney, "Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained Gaussian HMM: A comparison for offline handwriting recognition," *Proceedings - International Conference on Image Processing, ICIP*, pp. 3541–3544, 2011, doi: 10.1109/ICIP.2011.6116480.
[29] M. Kozielski, P. Doetsch, H. N.-2013 12th International, and undefined 2013, "Improvements in rwth's system for off-line handwriting recognition," *ieeexplore.ieee.orgM Kozielski, P Doetsch, H Ney2013 12th International Conference on Document Analysis and, 2013•ieeexplore.ieee.org*, pp. 935–939, 2013, doi: 10.1109/ICDAR.2013.190.
[30] A. Vinciarelli and J. Luettin, "A new normalization technique for cursive handwritten words," *Pattern Recognit Lett*, vol. 22, no. 9, pp. 1043–1050, Jul. 2001, doi: 10.1016/S0167-8655(01)00042-3.
[31] M. Kozielski, P. Doetsch, and H. Ney, "Improvements in RWTH's system for off-line handwriting recognition," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 935–939, 2013, doi: 10.1109/ICDAR.2013.190.
[32] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, "Improved handwritten digit recognition using convolutional neural networks (Cnn)," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–18, Jun. 2020, doi: 10.3390/s20123344.
[33] M. B. Bora, D. Daimary, K. Amitab, and D. Kandar, "Handwritten Character Recognition from Images using CNN-ECOC," *Procedia Comput Sci*, vol. 167, pp. 2403–2409, Jan. 2020, doi: 10.1016/J.PROCS.2020.03.293.
[34] A. Poznanski and L. Wolf, "CNN-N-Gram for Handwriting Word Recognition," 2016.
[35] T. Bluche and R. Messina, "Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 1, pp. 646–651, Jul. 2017, doi: 10.1109/ICDAR.2017.111.
[36] J. Wang and X. Hu, "Gated Recurrent Convolution Neural Network for OCR," *Adv Neural Inf Process Syst*, vol. 30, 2017.
[37] H. Nisa, "Off-line Handwritten Text Recognition with Cross-outs and Corrections Using Deep Learning," 2023.
[38] O. Tüselmann and G. A. Fink, "Neural models for semantic analysis of handwritten document images," *International Journal on Document Analysis and Recognition*, 2024, doi: 10.1007/s10032-024-00477-8.
[39] L. Kang, P. Riba, M. Rusiñol, A. Fornés, and M. Villegas, "Pay Attention to What You Read: Non-recurrent Handwritten Text-Line Recognition".
[40] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition," Jun. 2014, doi: 10.48550/arxiv.1406.2227.
[41] S. I. Google, "Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models," *Adv Neural Inf Process Syst*, vol. 30, 2017.
[42] P. Dreuw, P. Doetsch, C. Plahl, and H. Ney, "Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained Gaussian HMM: A comparison for offline handwriting recognition," *Proceedings - International Conference on Image Processing, ICIP*, pp. 3541–3544, 2011, doi: 10.1109/ICIP.2011.6116480.

پیوست

الف) شبه کد الگوریتم پیشنهادی دوم به صورت زیر می‌باشد:

Data: Handwritten English words
Result: Recognition of handwritten English words

Initialization:
- Load handwritten English word images
- Preprocess the images (resize, deslant, augment)

Convolutional Block:
- Repeat 3 times:
  - For each repetition:
    - Convolution with 3x3 window and 32 filters
    - ReLU activation

Additional Convolutional Layers:
- Convolution with 3x3 window and 256 filters
- ReLU activation
- Batch renormalization
- Gated convolution with 3x3 window and 256 filters
- Convolution with 3x3 window and 256 filters
- ReLU activation
- Batch renormalization

- 4 Maxpooling (same as previous Maxpooling operations)

Recurrent Layers (BLSTM):
- Bidirectional LSTM with 256*2 units
- Dropout with rate 0.4
- Bidirectional LSTM with 256*2 units
- Dropout with rate 0.2

Dense and Softmax Layer:
- Fully connected layer
- Softmax activation

Output:
- Recognized handwritten English words

- Batch renormalization
- Gated convolution with 3x3 window and 32 filters

- Maxpooling with kernel size (1, 2)

- Repeat 3 times:
  - For each repetition:
    - Convolution with 3x3 window and 64 filters
    - ReLU activation
    - Batch renormalization
    - Gated convolution with 3x3 window and 64 filters
    - Maxpooling with kernel size (1, 2)

- Repeat 3 times:
  - For each repetition:
    - Convolution with 3x3 window and 128 filters
    - ReLU activation
    - Batch renormalization
    - Gated convolution with 3x3 window and 128 filters
    - Maxpooling with kernel size (1, 2)
    - Maxpooling with kernel size (2, 1)

And also:
- Maxpooling with kernel size (1, 2)
- Maxpooling with kernel size (2, 1)

- Combine the outputs of the two Maxpooling operations

ب) تشریح لایه به لایه معماری

جدول ۴: تشریح لایه به لایه روش پیشنهادی

شماره لایه	لایه (نوع)	سایز خروجی	تعداد فیلترها	ابعاد کرنل	گام(stride)	نوع پدینگ	جزئیات
۱	ورودی	۱۲۸*۳۲*۱	-	-	-	-	-
۲	کانولوشن	۶۴*۱۶*۳۲	۳۲	۳*۳	۲*۲	same	مقدار دهی اولیه کرنل: he_uniform
	نرمال‌سازی دسته‌ای	۶۴*۱۶*۳۲	-	-	-	-	Renorm=True
۳	بلوک دروازه‌ای	۶۴*۱۶*۳۲	۳۲	۳*۳	-	same	-
۴	کانولوشن	۶۴*۱۶*۳۲	۳۲	۳*۳	۱*۱	same	مقداردهی اولیه کرنل: he_uniform
	نرمال‌سازی دسته‌ای	۶۴*۱۶*۳۲	-	-	-	-	Renorm=True
۵	بلوک دروازه‌ای	۶۴*۱۶*۳۲	۳۲	۳*۳	-	same	-
۶	کانولوشن	۶۴*۱۶*۳۲	۳۲	۳*۳	۱*۱	same	مقداردهی اولیه کرنل: he_uniform
	نرمال‌سازی دسته‌ای	۶۴*۱۶*۳۲	-	-	-	-	Renorm=True
۷	بلوک دروازه‌ای	۶۴*۱۶*۳۲	۳۲	۳*۳	-	same	Kernel Constraint: MaxNorm(4, [0, 1, 2])
۸	پولینگ	۶۴*۸*۳۲	-	-	۱*۲	valid	-
۹	کانولوشن	۶۴*۸*۶۴	۶۴	۳*۳	۱*۱	Same	مقداردهی اولیه کرنل: he_uniform
	نرمال‌سازی دسته‌ای	۶۴*۸*۶۴	-	-	-	-	Renorm=True
۱۰	بلوک دروازه‌ای	۶۴*۸*۶۴	۶۴	۳*۳	-	same	Kernel Constraint: MaxNorm(4, [0, 1, 2])
۱۱	کانولوشن	۶۴*۸*۶۴	۶۴	۳*۳	۱*۱	same	he_uniformمقداردهی اولیه کرنل:
	نرمال‌سازی دسته‌ای	۶۴*۸*۶۴	-	-	-	-	Renorm=True

پاورقی‌ها:

[1] Convolutional Recurrent Neural Network (CRNN)
[2] Gated Convolutional Neural Network (CNN)
[3] Data Augmentation
[4] Handwritten recognition
[5] Intelligent Character Recognition (ICR)
[6] Online Handwriting Recognition
[7] Offline Handwriting Recognition
[8] Desloping
[9] Deslanting
[10] Hidden Markov Models (HMMs)
[11] Elastic Distortions
[12] Long Short Term Memory (LSTM)
[13] Forget gate
[14] Time-steps
[15] Gated Linear Units
[16] Non-deterministic gates
[17] Scaling
[18] Preprocessing
[19] Padding
[20] Attention
[21] Multi-layer perceptron
[22] Classifier
[23] Otsu thresholding
[24] Optimizer
[25] Adam
[26] Error Correcting Output Code (ECOC)
[27] Dropout
[28] ReLU
[29] Dense layer
[30] Fully-Connected layer
[31] Feed-forward
[32] Bezier curve
[33] Multi-lingual word recognition
[34] Cascaded Script Identification
[35] Transfer Learning
[36] Dynamic Programming
[37] Multi-head self-attention
[38] Symbols
[39] Training Process
[40] Input Data
[41] Gated Hybrid Model
[42] Character Error Rate (CER)
[43] Word Error Rate (WER)
[44] Mini-batching
[45] Affine transformation
[46] Fully Connected OCR
[47] Batch normalization w.t.t. renormalization
[48] Gated Convolutional Optical Character Recognition (GCOCR)
[49] Pointwise multiplication
[50] Mini-block
[51] Feature map
[52] He uniform
[53] Blank symbol
[54] Character Error Rate (CER)
[55] Word Error Rate (WER)
[56] TensorFlow
[57] Learning rate
[58] Reduce learning rate on plateau
[59] Early stopping