

# تشخیص و فیلترینگ هوشمند تصاویر نامتعارف

## به کمک شبکه‌های عصبی عمیق



علی قنبری سرخی\*، منصور فاتح و حمید حسن پور

دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود، شاهرود، ایران

### چکیده

با پیشرفت روزافزون اینترنت و رسانه‌های تحت وب، توزیع و اشتراک منابع اطلاعاتی نظیر تصویر در حال افزایش است. اشتراک این منابع علاوه بر مزایای بسیار، خطرات و مشکلاتی نظیر دسترسی به تصاویر نامتعارف دارد که به نوبه خود تهدیدی برای فرهنگ جوامع مختلف، به خصوص نوجوانان و جوانان است. امروزه بسیاری از افراد، عضو سایت‌های اجتماعی از جمله اینستاگرام و فیسبوک هستند. به دلیل عدم وجود فیلترینگ هوشمند مناسب، حتی وجود درصدی اندک از تصاویر نامتعارف، فیلتر شدن کلی سایت‌های اجتماعی را به همراه دارد که برای کاربران، احساس ناراضی‌تری را به ارمغان می‌آورد. به همین منظور، در این مقاله به تحلیل و بررسی روشی برای دسته‌بندی تصاویر نامتعارف و فیلترینگ هوشمند آن‌ها پرداخته شده است. یکی از مشکلات این نوع از سامانه‌ها، حجم بالای داده‌های موجود در شبکه‌های تحت وب و استخراج ویژگی‌های معنادار در این حجم از داده‌ها است. در این راستا، در این مقاله روشی جدید، بر پایه شبکه‌های عصبی عمیق به منظور تشخیص هوشمند تصاویر نامتعارف ارائه شده است. این نوع از شبکه‌ها، مفاهیم سطح بالا را از روی ویژگی‌های سطح پایین استخراج می‌کنند و با این استخراج مفاهیم، به دقت مناسبی در دسته‌بندی اطلاعات دست می‌یابند. در این پژوهش، معماری جدیدی برای شناسایی تصاویر نامتعارف پیشنهاد شده است. نتایج به دست آمده بر روی مجموعه داده به نسبت بزرگ آزمایش شده است. این آزمایش‌ها نشان می‌دهد که روش پیشنهادی دو درصد دقت بیشتری نسبت به روش‌های جدید مطرح شده در شناسایی تصاویر نامتعارف دارد.

واژگان کلیدی: فیلترینگ هوشمند، تصاویر نامتعارف، شبکه عصبی عمیق، شبکه عصبی کانولوشن.

## Intelligent Identifications and Filtering of Unconventional Images Based on Deep Neural Networks

Ali Ghanbari Sorkhi\*, Mansour Fatch & Hamid Hassanpour

Faculty of Computer Engineering and IT, Shahrood University of Technology, Shahrood, Iran

### Abstract

Currently vast improvement of internet access and significant growth of web based broadcasters have resulted in distribution and sharing of informative resources such as images worldwide. Although this kind of sharing may bring many advantages, there are certain risks such as access of kids to porn images which should not be neglected. In fact, access to these images can be a threat to the culture of any society where kids and adults are included. However, many of internet users are members of social websites including Facebook or Instagram and without an appropriate intelligent filtering system, presence of few unconventional images may result in total filtering of these websites causing unpleasant feeling of members. In this paper, an attempt was made to propose an approach for classification and intelligent filtering of unconventional images. One of the major issues on these occasions is the analysis of a large scale of data available in the websites which might

\* Corresponding author

\* نویسنده عهده‌دار مکاتبات

فصلنامه علمی



۵۵

سال ۱۳۹۷ شماره ۲ پیاپی ۳۶

be a very time consuming task. A deep neural network might be a good option to resolve this issue and provide a good accuracy in dealing with huge databases. In this research, a new architecture for identifying unconventional images is proposed. In the proposed approach, the new architecture is presented with a combination of AlexNet and LeNet architecture that uses convolutional, polling and fully-connected layers.

The activation function used in this architecture, is the Rectified Linear Unit (ReLU) function. The reason of using this activation function is the high speed of convergence in deep convolution networks and simplicity in implementation. The proposed architecture consists of several parts. The first two parts consist of convolutional layers, ReLUs and pooling. In this section, convolution is applied to the input image with different dimensions and filters. In the next section, the convolutional layer with ReLU is used without pooling. The next section, like the first two parts, includes convolutional layers, ReLU and pooling. Finally, the last three parts include the fully-connected layers with ReLU. The output of the last layer is the two classes, which specifies the degree of belonging of each input to the class of unconventional and conventional images. The results are tested on a large-scale dataset. These tests show that the proposed method is more accurate than the other methods recently developed for identifying unconventional images.

**Keywords:** Intelligent filtering system, unconventional images, deep neural network, conventional neural network

اینکه روش دقیقی برای محاسبه تعداد وبسایت‌های مرتبط به تصاویر نامتعارف و همچنین بازدیدکنندگان آن وجود ندارد، اما می‌توان تعداد این وبسایت‌ها را بسیار بالا تخمین زد.

(جدول ۱-): آمارهای مرتبط به هزینه و تعداد افراد در

دسترسی به اطلاعات هرزه‌نگاری<sup>۲</sup> [2]

(Table-1): Related statistics Pornography according to cost and number of people [2]

\$۳۰۷۵,۶۴		در هر ثانیه	هزینه انجام شده برای پورن	
۲۸۲۵۸			تعداد مشاهده کنندگان پورن	
۳۲۷			تعداد افراد جستجو کننده پورن	
۷۰٪ بازه سنی ۲۴-۱۸	۱/۳ خانم	در هر ماه	۴۰ میلیون نفر	تعداد افراد جستجوکننده پورن در کشور آمریکا
	۲/۳ آقا			

جدول (۱) نتایج تکان‌دهنده‌ای را از گزارش ارائه‌شده در مرجع [2] در این زمینه نشان می‌دهد. گزارش‌های مطرح‌شده در مراجع [3-6] به بررسی اثرات هرزه‌نگاری در جوامع پرداخته است و نتایج این گزارش نشان می‌دهد که ارتباط مستقیمی بین هرزه‌نگاری و جنایت یا رفتارهای پرخاشگرانه جنسی در جامعه وجود دارد.

استفاده از اینترنت به‌عنوان یک مرجع مهم برای دانش‌آموزان در آموزش و پرورش و استفاده متناوب جوانان و نوجوانان از این منبع اطلاعات، همواره خطراتی را برای

<sup>۲</sup>پورنوگرافی یا هرزه‌نگاری، تجسمی بی‌پرده از مسایل جنسی با هدف تحریک یا ارضاست که در قالب‌های مختلف از جمله کتاب، فیلم، اینترنت و مجله ارائه می‌شود. در این مقاله تمرکز بر روی تصاویر است.

## ۱- مقدمه

اشتراک منابع اطلاعاتی به مانند تصویر و توزیع آن در سطح جهانی با رشد سریع اینترنت و رسانه‌های اجتماعی تحت وب روزبه‌روز در حال افزایش است. اگرچه این پدیده مزایای بسیاری همچون توسعه اشتراک محتوا در زمینه‌های آموزشی، خبری و غیره را داشته، ولی نگرانی‌هایی را در این حوزه نیز ایجاد کرده است. از جمله این نگرانی‌ها به انتشار تصاویر ناخواسته، توهین‌آمیز و نامتعارف<sup>۱</sup> در میان وبسایت‌های اینترنتی می‌توان اشاره کرد که اهمیت فیلترکردن این تصاویر را نشان می‌دهد. همچنین بسیاری از سایت‌های اجتماعی درصدی اندک از تصاویر نامتعارف دارند. در صورت عدم وجود فیلترینگ هوشمند، دولت مجبور به فیلترکردن کل سایت است که با توجه به استفاده وسیع کاربران از این‌گونه سایت‌ها، احساس عدم‌رضایت در میان کاربران اینترنتی ایجاد می‌شود. درواقع به‌کمک فیلترینگ هوشمند دسترسی کاربران به بسیاری از سایت‌ها را می‌توان ممکن و تنها دسترسی به اطلاعات نامتعارف را محدود کرد که تصاویر نامتعارف یکی از مهم‌ترین این اطلاعات محسوب می‌شوند. ما در این مقاله روشی را برای تشخیص تصاویر نامتعارف ارائه می‌دهیم که در موضوع فیلترینگ هوشمند اهمیت دارد.

بر اساس آمار ارائه‌شده، در سال ۲۰۱۱ میلادی دوازده درصد از سایت‌های موجود در شبکه اینترنت حاوی تصاویر نامتعارف بودند [1]. بر اساس این آمار، تصاویر نامتعارف منتشرشده در اینترنت توسط ۴۲,۷ درصد از کاربران اینترنتی مشاهده شده است و از این تعداد ۳۴ درصد از کاربران علاقه‌ای به دیدن این تصاویر نداشته‌اند. با وجود

<sup>1</sup> Pornography

در بخش دوم مقاله به معرفی کارهای انجام شده در این زمینه پرداخته می‌شود. در بخش سوم شبکه‌های عصبی عمیق و معماری شبکه پیشنهادی بیان می‌شود. در ادامه به معرفی مجموعه داده، آزمایش‌های انجام شده و همچنین تحلیل نتایج پرداخته می‌شود. در نهایت نیز نتیجه‌گیری از این پژوهش بیان می‌شود.

## ۲- کارهای انجام شده

در سال‌های اخیر پژوهش‌های بسیاری بر روی تشخیص تصاویر نامتعارف انجام شده است. در این پژوهش‌ها، روش‌های فیلترینگ تصاویر نامتعارف بر پایه محتوا و فیلترینگ هوشمند، جایگاه ویژه‌ای دارد. در این نوع از فیلترینگ، به‌طور معمول از طبقه‌بندی آماری استفاده می‌شود. در این طبقه‌بندی، تصاویر به دو دسته متعارف و نامتعارف تقسیم می‌شوند. به‌صورت کلی پژوهش‌های انجام شده برای شناسایی تصاویر نامتعارف به چهار دسته بر اساس ساختار بدن انسان، بازیابی تصویر، کلمات بصری<sup>۱</sup> و ویژگی‌های مبتنی بر پوست تقسیم می‌شود. در ادامه کارهای انجام شده با توجه به این تقسیم‌بندی معرفی و تشریح شده است.

روش‌های مبتنی بر ساختار بدن انسان از سال ۱۹۹۶ با پژوهش‌های انجام شده در مراجع [12, 13] مطرح شده است. در این نوع از روش‌ها، نواحی مرتبط با رنگ پوست تشخیص و در گام بعدی با توجه به اطلاعات ساختار بدن و نواحی استخراجی از رنگ پوست، قسمت‌های مرتبط با بدن انسان شناسایی و در نهایت با توجه به نواحی استخراج شده از بدن، تصاویر نامتعارف تشخیص داده می‌شود. با توجه به غیرصلب<sup>۲</sup> بودن بدن انسان و وجود حالت‌های مختلف از بدن در تصویر، به‌دست آوردن یک حالت جامع از مدل بدن انسان کار دشواری است. تاکنون مدل خوبی برای ساختار بدن انسان ارائه نشده است. از این رو، روش‌های مبتنی بر ساختار بدن انسان دقت کمی دارند [14, 15].

هدف از روش‌های مبتنی بر بازیابی<sup>۳</sup> تصویر، به‌دست آوردن بهترین انطباق تصویری برای شناسایی زیرطبقه‌های مجموعه داده است [16]. در این روش‌ها، تصاویر نامتعارف منطبق با تصویر ورودی از روی پایگاه داده استخراج می‌شود.

جامعه دارد و این افراد را در معرض خطر تصاویر نامتعارف قرار می‌دهد. با توجه به موارد ذکر شده، به‌طور معمول محدودیت‌های قانونی برای دسترسی به تصاویر نامتعارف برای افراد زیر سن قانونی اعمال می‌شود [7, 8]. با توجه به گسترده بودن این اطلاعات اعمال محدودیت همه مشکلات در این زمینه را نمی‌تواند رفع کند. در همین راستا، در سال‌های اخیر پژوهش‌های بسیاری در زمینه‌های پردازش تصویر برای طراحی سامانه‌ای به‌منظور فیلترینگ هوشمند، این نوع از تصاویر انجام شده است. با توجه به ماهیت تصاویر و پیچیدگی‌های موجود در ساختار تصویر، تشخیص و جداسازی تصاویر نامتعارف به‌سادگی امکان‌پذیر نیست.

در بسیاری از کارهای انجام شده، از رنگ پوست و ویژگی‌های مرتبط با آن برای تشخیص تصاویر نامتعارف استفاده شده است. تشخیص پیکسل‌های مرتبط با رنگ پوست به‌عنوان نخستین گام در شناسایی این نوع از تصاویر بوده است؛ اما به‌اندازه کافی برای شناسایی تصاویر نامتعارف مورد اطمینان نیستند [9]. در سامانه‌ای که برای تشخیص این نوع از تصاویر در گوگل راه‌اندازی شد، سرعت پردازش و دقت عمل، به‌خصوص برای حجم بالای داده‌ها مورد توجه است. در سامانه یادشده طبقه‌بندی بر پایه ۲۷ ویژگی استخراج شده از پوست و طبقه‌بندی ماشین بردار پشتیبان استفاده شده است؛ ولی با توجه به حجم بالای این نوع از داده‌ها، همچنان این پژوهش نیازمند پژوهش‌های بیشتر است [10].

با توجه به موارد بیان شده، طراحی سامانه‌ای قدرتمند برای آنالیز تصاویر و فیلترینگ هوشمند آن‌ها بسیار مهم و حیاتی است. از طرفی در سال‌های اخیر استفاده از شبکه‌های عصبی عمیق برای شناسایی و تشخیص تصاویر در مجموعه داده بزرگ مورد استفاده قرار گرفته و دارای عملکرد مناسبی است [11]. در این راستا در این مقاله، سامانه‌ای جامع بر پایه یادگیری شبکه‌های عصبی عمیق با ساختار نوین برای فیلترینگ هوشمند تصاویر نامتعارف معرفی شده است. این ساختار نوین در بخش ۲-۳ به‌طور کامل شرح داده شده است. روش معرفی شده در این مقاله بر روی تصاویر رنگی استخراجی از صفحات وب با کیفیت‌های متفاوت مورد ارزیابی قرار گرفته است. همچنین نتایج روش‌های جدید بر روی تصاویر رنگی گزارش شده است؛ ولی غالب این نتایج را بر روی تصاویر خاکستری نیز مورد ارزیابی می‌توان قرار داد. نتایج نشان‌دهنده عملکرد بهتر این روش نسبت به روش‌های دیگر است.

1 Image Retrieval

2 Visual words

3 Non-rigid

4 Retrieval

در صورتی که تعداد تصاویر نامتعارف انطباقی بیشتر از حد آستانه تعیینی باشد، تصویر ورودی به عنوان تصویر نامتعارف معرفی می‌شود. شناسایی در این روش‌ها، تا حد زیادی وابسته به پایگاه داده است. با توجه به فراوانی تصاویر متعارف و نامتعارف در شکل‌های مختلف، ساخت پایگاه داده کامل بسیار دشوار است. برای بهبود دقت شناسایی در این گونه روش‌ها، نیاز به تعداد بالای نمونه‌ها است که زمان شناسایی را در مرحله آزمون کاهش می‌دهد [15].

در سال‌های اخیر شناسایی تصاویر مبتنی بر کلمات بصری بسیار مورد توجه پژوهش‌گران قرار گرفته و این روش، محبوبیت زیادی در طبقه‌بندی اشیاء به دست آورده است [17]. با توجه به تعریف ارائه شده در [18] کلمات بصری به قطعات کوچکی از تصویر اشاره می‌کنند که نوعی از اطلاعات مرتبط با ویژگی‌هایی مانند رنگ، شیء یا بافت را دارد. پژوهش‌گران روشی برای تحلیل تصویر با الهام‌گیری از تحلیل محتوای متن ارائه کرده‌اند. در این روش، یک تصویر به عنوان ترکیبی از کلمات در نظر گرفته و روش تحلیل متنی مورد استفاده در تفسیر معنایی<sup>۱</sup> برای شناسایی تصاویر نامتعارف استفاده می‌شود [19]. در این روش‌ها کلمات بصری برای توصیف محتوای معنایی تصویر استخراج می‌شوند؛ سپس مدل کیسه‌ای از کلمات (BoW<sup>۲</sup>) برای شناسایی تصاویر نامتعارف معرفی می‌شود [20]. کلمات بصری و اطلاعات معنایی از برجسب‌های مرتبط<sup>۳</sup>، برای تخمین امتیاز ارتباط بین مدل BoW جهت شناسایی برجسب‌های موجود در تصاویر استفاده [21, 22] و تصاویر نامتعارف در مرجع [23] استفاده شده است. در پژوهش [24] توصیف‌گر SIFT<sup>۴</sup> برای تولید واژگان<sup>۵</sup> بصری استفاده شده و برای شناسایی تصاویر نامتعارف از طبقه‌بند ماشین بردار پشتیبان استفاده شده است. در مرجع [25] از اطلاعات متن و کیسه‌ای از کلمات بصری (BoVW)<sup>۶</sup> برای شناسایی تصاویر نامتعارف استفاده شده است. در این پژوهش به جای توصیف‌گر SIFT از روش SURF<sup>۷</sup> استفاده شده است [26]. در روش‌های مبتنی بر ویژگی‌های پوستی، شناسایی تصاویر به مانند یک مسئله طبقه‌بندی است [27]. در این

روش‌ها، ویژگی‌های رنگی و بافتی از پوست استخراج می‌شوند و در گام بعدی از طبقه‌بندهای الگو برای شناسایی و طبقه‌بندی تصاویر استفاده می‌شوند [28-30]. بسیاری از کارهای انجام شده در سال‌های اخیر مرتبط به این روش بوده است. کارهای انجام شده در این زمینه، به سه دسته کلی مبتنی بر اطلاعات<sup>۸</sup> رنگ، اطلاعات شکل<sup>۹</sup> و توصیف‌گرهای ویژگی‌های محلی تقسیم می‌شوند. روش‌های مبتنی بر ویژگی‌های رنگی پیکسل، به‌طور معمول برای شناسایی نواحی پوست استفاده می‌شوند. در مرجع [31] بیش از دویست هزار تصویر از سایت‌های متعارف و نامتعارف انتخاب شده است که مقدار فام<sup>۱۰</sup> در این دو دسته از تصاویر بسیار متفاوت است (شکل ۱). هر پیکسل از تصویر با توجه به مدل رنگ می‌تواند برای شناسایی پیکسل‌های پوستی و غیرپوستی استفاده شود. این روش‌ها به‌طور معمول توسط طبقه‌بندهای ماشین بردار پشتیبان [32, 33]، شبکه‌های عصبی پرسپترون [27] و درخت تصمیم [15] دسته‌بندی می‌شوند.

دسته دوم روش‌های فیلترینگ هوشمند مبتنی بر پوست، بر اطلاعات شکل هستند. این روش‌ها بر تشخیص نواحی پوستی تکیه دارند. همچنین در این روش‌ها از تعدادی ویژگی‌های مبتنی بر توصیف‌گر نواحی تصویر استفاده شده است. ویژگی‌های مرتبط به شکل را در پنج دسته، ویژگی‌های مبتنی بر لبه<sup>۱۱</sup> [34, 35]، گشتاورها [36] [37]، محدودیت‌های هندسی<sup>۱۲</sup> [38]، قطعه‌بندی رنگ<sup>۱۳</sup> [39] و ویژگی‌های MPEG<sup>۷</sup> [40, 41] می‌توان دسته‌بندی کرد.

دسته سوم روش‌های فیلترینگ هوشمند مبتنی بر پوست، از بردار ویژگی برای توصیف یک ناحیه از تصویر استفاده می‌کنند. در این روش‌ها نقاط کلیدی<sup>۱۴</sup> با نمونه‌گیری تنک<sup>۱۵</sup> از تصویر انتخاب می‌شوند و از این نقاط به عنوان ویژگی برای شناسایی تصاویر مشابه استفاده شده است. در کارهای انجام شده توسط این روش به SIFT [25, 42] و PLSA<sup>۱۶</sup> [43, 44] می‌توان اشاره کرد.

8 Information

9 Shape

10 Hue

11 Contour

12 Geometric constraints

13 Color segments

14 Key point

15 Sparse

16 Probabilistic Latent Semantic Analysis

1 Semantic annotation

2 Bag-of-Word (BoW)

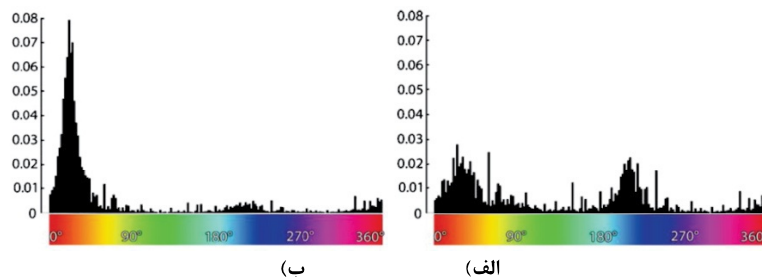
3 Associated tags

4 Scale Invariant Feature Transform

5 Visual vocabulary

6 Bag-of-Visual-Word (BoVW)

7 Speeded up robust features



(شکل-1): مقایسه بین نتایج به دست آمده از توزیع مقدار فام (Hue) در تصاویر استخراج شده از وبسایت‌های متعارف و نامتعارف [31]. مقدار نرمال شده فام با زاویه بین ۰ تا ۳۶۰ درجه با توجه به تعریف فضای رنگی HSV محاسبه شده است. الف) تصاویر متعارف، ب) تصاویر نامتعارف

(Figure-1): The comparison between the results obtained from the distribution of the amount of hue in images extracted from conventional and unconventional websites [31], the normalized value of the Hue by an angle between 0° and 360° degrees is calculated according to the definition of the HSV color space. a) Conventional images, b) unconventional images

چالش اصلی در این نوع از روش‌ها، استخراج ویژگی مؤثر و مفید برای جداسازی تصاویر متعارف از نامتعارف است [17]. در مرحله نخست، از تصاویر ویژگی استخراج می‌شود و در گام بعدی طبقه‌بندی انجام می‌شود که عملکرد طبقه‌بند به مرحله استخراج ویژگی بسیار وابسته است. در این مقاله از یادگیری ماشینی با شبکه‌های عصبی عمیق کانولوشن برای استخراج ویژگی و طبقه‌بندی هم‌زمان استفاده شده است.

ناحیه ادراکی محلی<sup>۴</sup> و نمونه‌گیری مکانی کاهشی<sup>۵</sup> احتمال وقوع بیش‌برازش کاهش می‌یابد. در برخی از کاربردها، استخراج ویژگی‌های معنادار دشوار است یا ویژگی‌های معنادار استخراج شده، بازنمایی مناسبی از داده‌ها ندارند. در این راستا، استخراج ویژگی‌ها از طریق یادگیری مهم به نظر می‌رسد. یادگیری عمیق، بر اساس یادگیری چند سطحی از بازنمایی‌های مختلف بر روی یک ساختار سلسله‌مراتبی از ویژگی‌ها است. در یادگیری عمیق، مفاهیم سطح بالا از روی ویژگی‌های سطح پایین تعریف می‌شوند و مفاهیم سطح پایین نیز به تعریف مفاهیم سطح بالاتر کمک می‌کنند [45]. هدف یادگیری عمیق، استخراج هوشمندانه ویژگی‌ها طی چندین مرحله یادگیری است. یکی از پژوهش‌های مهم برای تشخیص مفاهیم سطح بالاتر توسط تیم گوگل در سال ۲۰۱۲ انجام شده است [46].

در این مقاله، از شبکه‌های عمیق برای استخراج ویژگی‌های مناسب و طبقه‌بندی تصاویر نامتعارف استفاده شده است که روشی نوین و کارا در این زمینه است. شبکه‌های عمیق انواع متفاوتی دارند که از مهم‌ترین آنها به شبکه‌های کانولوشن، شبکه باور عمیق<sup>۶</sup>، شبکه عصبی حافظه کوتاه مدت ماندگار<sup>۷</sup> و غیره می‌توان اشاره کرد. یکی از شبکه‌های عمیق استفاده شده در پردازش تصویر شبکه‌های کانولوشن است که در این پژوهش نیز از این نوع

۳- شبکه عصبی عمیق

امروزه شبکه‌های عصبی به دلیل قابلیت مناسب در یادگیری ویژگی‌های ترکیبی و نوین، در دسته‌بندی استفاده می‌شوند. شبکه عصبی عمیق، مدلی از شبکه عصبی است که برای یادگیری تبدیل غیرخطی روی داده‌ها مناسب است. تفاوت اساسی مدل‌های شبکه عمیق با شبکه عصبی معمولی، حفظ قابلیت بازسازی داده در هر لایه است. به عبارت دیگر شبکه عمیق، با مدل کردن توزیع داده‌ها در هر لایه، اطلاعات فضای ویژگی را حفظ می‌کند. این شبکه، قابلیت تعمیم‌پذیری بیشتری روی داده‌های آزمون دارد و در آن بیش‌برازش<sup>۱</sup> مدل رخ نمی‌دهد. در واقع در شبکه‌های عصبی معمولی وجود لایه‌هایی با اتصال کامل سبب اتلاف<sup>۲</sup> شده و از طرفی تعداد زیاد پارامترها در این نوع از شبکه‌ها، سرعت بیش‌برازش را افزایش می‌دهند؛ ولی در شبکه‌های عصبی عمیق به دلیل استفاده از روش‌های اشتراک وزن<sup>۳</sup>،

<sup>4</sup> Local receptor field

<sup>5</sup> Spatial down sampling

<sup>6</sup> Deep belief Network

<sup>7</sup> Long Short-Term Memory

<sup>1</sup> Overfitting

<sup>2</sup> Wasteful

<sup>3</sup> Weight Sharing

شبکه استفاده شده است. در ادامه به شرح این نوع از شبکه‌ها پرداخته می‌شود.

### ۱-۳- شبکه‌های عصبی کانولوشن

ورودی‌های شبکه‌های عصبی کانولوشن، تصاویر هستند. در ساختار این نوع از شبکه‌ها از روش اصلی شامل اشتراک وزن، ناحیه درک محلی و نمونه‌گیری مکانی کاهشی استفاده می‌شود. این شبکه تا حدی به مدل دستگاه بینایی بیولوژیکی شبیه است [47]. از لحاظ مفهومی، شبکه‌های عصبی کانولوشن را سامانه‌ای از چندین تشخیص‌دهنده سلسله‌مراتبی متصل به می‌توان هم در نظر گرفت.

در لایه نخست این شبکه، به‌طورمعمول تطبیق الگو<sup>۱</sup> غیرخطی با وضوح مکانی مناسب انجام می‌گیرد، و ویژگی‌های اصلی داده ورودی استخراج می‌شوند. از آنجایی که از یک تابع کرنل با یک شعاع همسایگی برای هر نقطه استفاده شده، تطبیق الگو غیرخطی است. لایه بعدی وجود ترکیب مکانی ویژگی‌های قبلی را تشخیص می‌دهد. گاهی عمل نمونه‌برداری کاهشی نیز با این شبکه همراه است. در این صورت، لایه‌های بعدی عمل شناسایی الگو را در مقیاس مکانی وسیع‌تر و با وضوح پایین‌تر انجام می‌دهند. با استفاده از خاصیت نمونه‌برداری کاهشی در شبکه عصبی کانولوشن، با تعداد محدودی وزن، ورودی‌های بزرگی را می‌توان پردازش کرد.

لایه کانولوشن (ConV) هسته اصلی تشکیل‌دهنده شبکه عصبی کانولوشن است. این لایه شامل نورون‌هایی در قالب توده‌های سه‌بعدی با عرض، ارتفاع و عمق مشخص است. تعداد نورون‌های هر توده، متناسب با فضای تصویر ورودی تعیین می‌شوند. خروجی این لایه، یک توده با توزیع مکانی را تشکیل می‌دهد که توسط لایه کانولوشنی بعدی مورد پردازش قرار می‌گیرد و در نهایت به لایه خروجی منتقل می‌شود. در واقع لایه بعدی توسط لغزاندن فیلتری بر روی توده لایه فعلی به‌دست می‌آید.

استفاده از نمونه‌برداری کاهشی یا نظرسنجی<sup>۲</sup> در بین لایه‌ها، وضوح لایه‌های بعدی را کاهش می‌دهد. لایه نظرسنجی به‌طورمعمول بین چندین لایه کانولوشنی قرار می‌گیرد. این لایه موجب کاهش اندازه تصویر و در نتیجه کاهش محاسبات در داخل شبکه و کنترل بیش‌برازش<sup>۳</sup> می‌شود. لایه نظرسنجی به‌صورت مستقل بر روی هر برش

عمقی از توده ورودی عمل کرده و آن را از لحاظ مکانی تغییر اندازه<sup>۴</sup> می‌دهد. از طرفی باید اشاره کرد که این لایه خاصیت تغییرناپذیری انتقال مکانی<sup>۵</sup> دارد؛ در واقع نسبت به جابه‌جایی مقاوم است. تابع‌های نظرسنجی بسیاری برای شبکه‌های کانولوشن مانند تابع بیشینه‌گیری و متوسط‌گیری معرفی شده است.

لایه تابع فعال‌ساز به‌طورمعمول قبل از لایه نظرسنجی در این نوع از شبکه‌ها قرار دارد. در این لایه بر روی تک تک نورون‌ها یک تابع فعال‌ساز اعمال می‌شود. همانطور که اشاره شده در اینجا از تابع فعال‌ساز با عنوان واحد خطی اصلاح‌کننده (ReLU<sup>۶</sup>) استفاده شده است. به‌طورمعمول پس از این لایه، لایه نرمال‌سازی قرار می‌گیرد. شرح مفصل این لایه در بخش ۳-۲ آمده است. لایه بعدی استفاده شده در این نوع از شبکه‌ها، لایه تمام متصل (FC<sup>۷</sup>) است. نورون‌هایی که در این لایه قرار دارند با تمام نورون‌های موجود در لایه قبلی (مشابه شبکه‌های عصبی معمولی) ارتباط دارند. بنابراین از ضرب ماتریسی و سپس جمع نتیجه حاصله با بایاس برای محاسبه تمامی نورون‌ها می‌توان استفاده کرد.

با توجه به توضحات ارائه شده، در ساختار یک شبکه عصبی کانولوشن، لایه‌های ورودی، کانولوشن، تابع فعال‌ساز، نرمال‌سازی، نظرسنجی و تمام متصل وجود دارند. در این نوع از شبکه‌ها باید تعداد لایه‌های کانولوشنی، لایه‌های نظرسنجی، اندازه فیلترها، تعداد نورون‌ها و نحوه اتصالات بین لایه‌ها مشخص شود. برای آموزش شبکه کانولوشنی همانند شبکه‌های عصبی معمولی از روش انتشار خطا رو به عقب استفاده می‌شود [11].

در LeNet به‌ترتیب از لایه‌های کانولوشن و نظرسنجی، کانولوشن و نظرسنجی، و سه لایه تمام متصل استفاده شده است. در شبکه AlexNet به‌ترتیب از لایه‌های کانولوشن، نظرسنجی، دو کانولوشن، دو نظرسنجی، و دو لایه تمام‌متصل استفاده شده است. در واقع در شبکه‌های قدیمی‌تر، قرارگرفتن لایه نظرسنجی بعد از کانولوشن امری رایج بوده است. در این مقاله از معماری برپایه AlexNet استفاده شده که در ادامه معماری پیشنهادی تشریح شده است.

<sup>4</sup> Resize

<sup>5</sup> Translation invariance

<sup>6</sup> Rectified Linear Unit (ReLU)

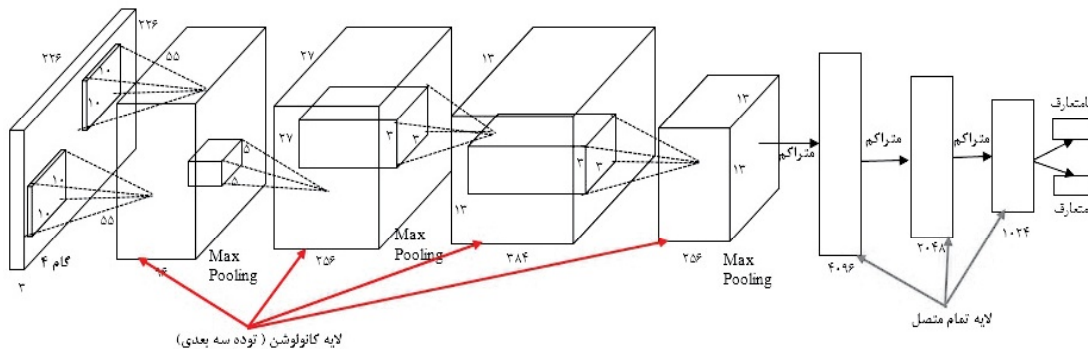
<sup>7</sup> Fully Connected (FC)

<sup>1</sup> Template Matching

<sup>2</sup> Pooling

<sup>3</sup> Overfitting





(شکل-۲): نمای کلی از معماری پیشنهادی برای شبکه عصبی کانولوشن  
(Figure-2): An overview of the proposed architecture for the convolutional network

## ۲-۳- معماری پیشنهادی برای شبکه‌های عصبی کانولوشن

آن،  $3 \times 1 \times 1$  به همراه گام ۱ چهار است. بلوک‌ها با اندازه  $3 \times 1 \times 1$  پیکسل به یک بردار ستونی با اندازه  $3 \times 1 \times 1 = 3$  تبدیل می‌شوند. برای محاسبه اندازه مکانی) توده خروجی از رابطه (۲) استفاده می‌شود:

$$\frac{W-F}{S} + 1 \quad (2)$$

در رابطه (۲)،  $W$  اندازه تصویر ورودی،  $F$  اندازه ناحیه ادراکی نورون‌های لایه کانولوشن و  $S$  اندازه گام هستند. در این پژوهش،  $W$ ،  $F$  و  $S$  به ترتیب ۲۲۶، ۱۰ و ۴ هستند. با تکرار این عمل بر روی ورودی، سرانجام ماتریس خروجی با اندازه  $[3025 \times 300]$  تولید می‌شود. هر ستون این ماتریس یک ناحیه ادراکی<sup>۲</sup> است و تعداد  $3025 = 55 \times 55$  ستون در کل وجود دارد. نواحی ادراکی با هم اشتراک دارند. از این رو، امکان تکرار چندین باره هم‌پوشانی<sup>۳</sup> هر عدد در توده خروجی وجود دارد. در این لایه ۹۶ کرنل یا فیلتر متفاوت در نظر گرفته شده است.

به همین ترتیب، وزن‌های شبکه کانولوشن به بردارهای سطری تبدیل می‌شوند. در لایه بعدی، ۹۶ فیلتر (مجموعه وزن‌ها) با اندازه  $3 \times 1 \times 1$  وجود دارد که تعداد وزن‌ها برابر با تعداد عناصر موجود در ناحیه ادراکی است. یعنی در ناحیه ادراکی با اندازه  $3 \times 1 \times 1$ ، تعداد وزن‌ها نیز همین مقدار خواهد بود. این عمل باعث ایجاد ماتریس جدید با اندازه  $[300 \times 96]$  خواهد شد.

طراحی معماری شبکه در عملکرد سامانه پیشنهادی بسیار تأثیرگذار است. در روش پیشنهادی، معماری جدید با ترکیب معماری AlexNet و LeNet ارائه شده که از لایه‌های کانولوشن، نظرسنجی و تمام‌متصل استفاده کرده است. بخش‌هایی از انتخاب هر لایه به مانند AlexNet است و از چند لایه کانولوشن پشت‌سرهم با تابع فعال‌ساز ReLU استفاده شده است. در این معماری، بخش‌هایی به مانند LeNet است و بعد از هر لایه کانولوشن یک لایه نظرسنجی آمده است. به صورت تجربی نشان داده شده است که این معماری عملکرد خوبی در شناسایی تصاویر نامتعارف دارد. شکل (۲) ساختار معماری پیشنهادی را نشان می‌دهد. معماری پیشنهادی، به صورت رابطه (۱) است. همان‌طور که در رابطه (۱) نشان داده شده، معماری پیشنهادی شامل چندین قسمت است. دو قسمت نخست شامل، لایه‌های کانولوشن، تابع فعال‌ساز ReLU و نظرسنجی (Pooling)، یک قسمت بعدی دارای لایه کانولوشن به همراه تابع فعال‌ساز ReLU، بعدی شامل لایه‌های کانولوشن، تابع فعال‌ساز ReLU و نظرسنجی (Pooling) و سه قسمت آخر شامل لایه تمام‌متصل با تابع فعال‌ساز ReLU هستند.

(۱)

$$\begin{aligned} input &\rightarrow [ConV \rightarrow ReLU \rightarrow Pooling] \times 2 \rightarrow \\ &[ConV \rightarrow ReLU] \rightarrow [ConV \rightarrow ReLU \rightarrow Pooling] \\ &\rightarrow [FC \rightarrow ReLU] \times 3 \rightarrow OutPut \end{aligned}$$

در شبکه پیشنهادی، اندازه تصویر ورودی  $226 \times 226 \times 3$  و اندازه فیلترهای استفاده شده در

<sup>1</sup>Stride  
<sup>2</sup>Receptive Field  
<sup>3</sup>Overlap

نتیجه عملیات کانولوشن (ضرب نقطه به نقطه) برابر با اجرای یک ضرب ماتریسی بزرگ است. این عملیات نتیجه ضرب نقطه‌ای بین تمام فیلترها و تمام نقاط نواحی ادراکی را تولید می‌کند. خروجی این عملیات در شبکه، ماتریسی با اندازه  $[30 \times 25 \times 96]$  است که نتیجه ضرب نقطه‌ای بین هر فیلتر در هر موقعیت است. نتیجه نهایی به صورت  $55 \times 55 \times 96$  تغییر شکل داده می‌شود.

لایه تابع فعال‌ساز، بر روی تک‌تک نورون‌ها، یک تابع فعال‌ساز را اعمال می‌کند. این تابع، اندازه توده مرحله قبل را تغییر نمی‌دهد. تابع فعال‌سازی ReLU با ورودی  $x$ ، در رابطه (۳) نشان داده شده است:

$$f(x) = \text{Max}(0, x) \quad (3)$$

استفاده از تابع فعال‌ساز ReLU، سرعت هم‌گرایی گرادبان نزولی تصادفی را نسبت به تابع سیگموئید و تانژانت هیپربولیک بهبود می‌دهد و پیاده‌سازی این روش نسبت به تابع‌های دیگر ساده‌تر است [11].

به‌طور معمول بر روی خروجی تابع فعال‌ساز، نرمال‌سازی داده‌ها انجام می‌شود [11]؛ اما تابع فعال‌ساز ReLU خواص مطلوبی<sup>۱</sup> دارد و نیازی به نرمال‌کردن ورودی برای اجتناب از اشباع‌شدن<sup>۲</sup> ندارد. با این وجود، برای تعمیم‌پذیری بیشتر این الگوریتم، از رابطه نرمال‌سازی (۴) استفاده شده است:

$$b_{x,y}^i = a_{x,y}^i / (k + \alpha \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} (a_{x,y}^j)^2)^\beta \quad (4)$$

در رابطه (۴)،  $a_{x,y}^i$  فعالیت<sup>۳</sup> نورون محاسبه‌شده توسط کرنل  $i$  در مکان  $(x,y)$  است و  $b_{x,y}^i$  فعالیت نرمال‌شده است.  $N$  تعداد کرنل‌ها در این لایه و  $k, n, \alpha, \beta$  مقادیر ابرپارامتر هستند که در مجموعه اعتبارسنجی استخراج می‌شوند. در این پژوهش،  $k=2, n=5, \alpha=10^{-4}, \beta=0.75$  در نظر گرفته شده‌اند. رایج‌ترین شکل استفاده از لایه نظرسنجی با فیلترهایی با اندازه  $2 \times 2$  به همراه گام<sup>۲</sup> است که هر برش عمقی در ورودی را با حذف دو عنصر از عرض و دو عنصر از ارتفاع کاهش داده و باعث حذف ۷۵٪ مقادیر موجود در آن برش

عمقی می‌شود. لایه نظرسنجی تصویر ورودی را در هر برش عمقی به صورت مستقل (یعنی هر برش بدون توجه به برش دیگر) از لحاظ مکانی کاهش<sup>۴</sup> می‌دهد. همان‌طور که در بخش قبل ذکر شد، تابع‌های نظرسنجی زیادی برای شبکه‌های کانولوشن معرفی شده که معروف‌ترین آنها تابع نظرسنجی بیشینه است. این تابع یک تابع کاهش نمونه غیرخطی است و از هر چهار عنصر، بیشینه مقدار را حفظ می‌کند.

لایه‌های بعدی به مانند یک شبکه عصبی پرسپترون دارای اتصالات تمام متصل هستند که هر لایه شامل  $4096, 2048$  و  $1024$  نرون می‌باشند، در نهایت لایه خروجی حاوی خروجی و امتیاز دسته‌ها است که در این مقاله لایه خروجی، امتیاز دو دسته متعارف و نامتعارف است.

در کاربرد فیلترینگ هوشمند تصاویر در فرهنگ‌های غربی تصاویر را به سه دسته اخلاقی، غیراخلاقی و شبه غیراخلاقی می‌توان دسته‌بندی کرد؛ اما با توجه به فرهنگ موجود در کشور ایران، تصاویر فقط به دو دسته متعارف (مطابق با فرهنگ) و غیرمتعارف (در تضاد با فرهنگ جامعه) دسته‌بندی می‌شوند که تصاویر شبه‌غیراخلاقی در دسته تصاویر نامتعارف قرار می‌گیرند.

## ۴- پیاده‌سازی و نتایج

در این بخش، به بیان مجموعه داده و مقایسه روش پیشنهادی با برخی از روش‌های انجام‌شده در شناسایی تصاویر نامتعارف پرداخته می‌شود. پیاده‌سازی و ارزیابی نتایج این مقاله در محیط متلب ۲۰۱۵ بر روی رایانه‌ای با پردازنده Core i7 2.67GHz با چهار هسته و حافظه جانبی چهار گیگا بایت انجام شده است.

### ۴-۱- مجموعه داده

مجموعه داده استفاده‌شده برای این مقاله از مرجع [42] بدست آمده است. تصاویر استفاده شده در این مجموعه، شامل دو دسته تصویر متعارف و نامتعارف هستند. در این مجموعه داده از ۱۸۳۳۵ تصویر، ۹۰۵۹ تصویر متعارف و ۹۲۹۵ تصویر نامتعارف از صفحات مختلف وب جمع‌آوری شده است. در تصاویر انتخاب‌شده بسیاری از حالات ممکن و قسمت‌های مختلف از هر دو دسته در نظر گرفته شده است.

<sup>4</sup>Downsample

<sup>1</sup> Desirable property

<sup>2</sup> Saturating.

<sup>3</sup> Activity



اعتبارسنجی، با دو معماری شبکه عصبی کانولوشن پیشنهادی و مرجع [11] را نشان می‌دهد. همان‌طور که از شکل (۳) مشخص است، شبکه کانولوشن پیشنهادی در مجموعه آموزش عملکرد بهتری داشته است. معماری پیشنهادی میزان خطای کمتری در شبکه کانولوشن نسبت به روش مرجع [11] در مجموعه آموزش و اعتبارسنجی از خود نشان داده است. مقایسه نتایج دو معماری در جدول (۲) جداگانه، نشان داده شده و برای مقایسه از معیار TP و FP [42] استفاده شده است. TP نشان‌دهنده نسبت تصاویر نامتعارف صحیح شناسایی شده به کل تصاویر نامتعارف و FP نشان‌دهنده نسبت تصاویر متعارف ناصحیح شناسایی شده به کل تصاویر متعارف است.

(جدول-۲): نتایج به دست آمده به صورت جداگانه توسط روش پیشنهادی و شبکه عصبی کانولوشن [11]

(Table-2): The obtained results are separated by a randomized method and a convolutional neural network [11]

		خروجی پیش‌بینی شده			
		نامتعارف (مثبت)		متعارف (منفی)	
		پیش‌بینی [۱۱]		پیش‌بینی [۱۱]	
		TP	FN	FP	TN
خروجی واقعی	نامتعارف (مثبت)	۹۱/۳۵	۸۹/۵۰	۸/۶۵	۱۰/۵
	متعارف (منفی)	۵/۱۱	۹/۲۱	۹۴/۸۹	۹۰/۷۹

با توجه به جدول (۲) نتایج به دست آمده از معماری پیشنهادی نسبت به معماری AlexNet در مرحله آزمون عملکرد بهتری دارد.

جدول (۳) معیارهای معرفی شده در بخش ارزیابی را بر روی این دو معماری نشان می‌دهد. نتایج جدول (۳) نشان می‌دهد که شبکه کانولوشن با معماری پیشنهادی عملکرد بهتری نسبت به معماری پیشنهاد شده در مرجع [11] دارد. در جدول (۳)، زمان شبکه در مرحله آموزش و آزمایش برای هر تصویر ارائه شده است. مقایسه زمان اجرای الگوریتم‌ها، نشان‌دهنده بهبود عملکرد زمانی شبکه کانولوشن پیشنهادی نسبت به روش مرجع [11] در مرحله آموزش و آزمایش است.

لازم به ذکر است که تنها از یکسان‌سازی اندازه تصاویر ورودی به عنوان پیش‌پردازش استفاده شده است و پیش‌پردازش یا پس‌پردازش دیگری بر روی تصاویر ورودی اعمال نشده است.

## ۲-۴- معیار ارزیابی

برای نشان دادن عملکرد روش پیشنهادی در شناسایی تصاویر نامتعارف از معیارهای کمی دقت<sup>۱</sup> (PR)، یادآوری<sup>۲</sup> (RE) و F1 استفاده می‌شود که این معیارها در مقاله‌های مختلفی برای ارزیابی روش‌های تشخیص تصاویر نامتعارف استفاده شده‌اند [51, 52]. معیار PR، مبتنی بر دقت پیش‌بینی دسته‌بند و مبین میزان اعتماد به خروجی دسته‌بند است. معیار RE، کارایی دسته‌بند را با توجه به تعداد رخداد یک دسته خاص نشان می‌دهد. معیار F1 ترکیبی از معیارهای دقت و یادآوری است، این معیارها در زیر معرفی شده‌اند:

$$PR = \frac{TP}{TP+FP} \quad (5)$$

$$RE = \frac{TP}{TP+FN} \quad (6)$$

$$F1 = \frac{2*RE*PR}{PR+RE} \quad (7)$$

در رابطه بالا منظور از TP تعداد تصاویری است که به درستی توسط طبقه‌بند به طبقه  $C_i$  انتساب یافته و FP یعنی تعداد تصاویر متعلق به سایر طبقه‌ها که به اشتباه توسط طبقه‌بند به طبقه  $C_i$  منسوب شده‌اند. FN تعداد تصاویر متعلق به طبقه  $C_i$  که توسط طبقه‌بند به سایر طبقه‌ها تعلق گرفته است. معیار TN تعداد تصاویری است که متعلق به طبقه  $C_i$  نبوده‌اند و توسط طبقه‌بند به این طبقه اختصاص داده نشده‌اند.

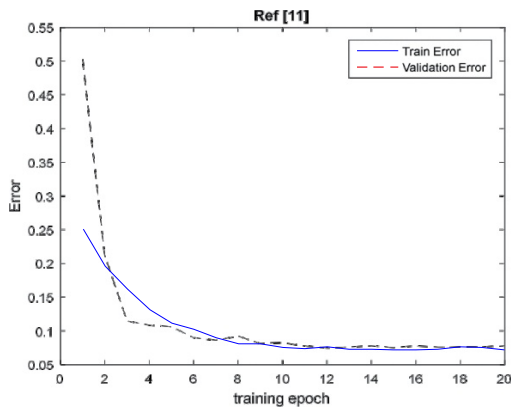
## ۳-۴- نتایج آزمایش‌ها

در مجموعه داده معرفی شده دوسوم داده‌ها برای آموزش و یک‌سوم برای آزمون در نظر گرفته شده است. در مجموعه‌ی آموزش یک‌سوم برای اعتبارسنجی در نظر گرفته شده است.

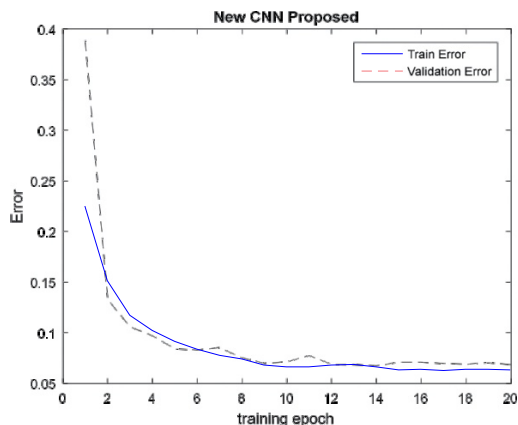
در گام نخست مقایسه‌ای بین معماری معروف AlexNet و معماری پیشنهادی انجام شده است. شکل (۳) نتایج به دست آمده از دسته‌بندی در مرحله آموزش و

<sup>1</sup> Precision

<sup>2</sup> Recall



(ب)



(الف)

(شکل-۳): نتایج مرحله آموزش و اعتبارسنجی در شبکه عصبی کانولوشن (خط پر و خط چین به ترتیب نشان دهنده مقدار خطای مجموعه آموزش و اعتبارسنجی است). الف) با معماری شبکه مرجع [11] ب) با معماری شبکه پیشنهادی (Figure-3): The results of the training phase and validation in the convolutional network (solid line and dashed line indicate the amount of error in the training and validation, respectively). a) Network architecture of reference [11] b) proposed network architectures

(جدول-۳): نتایج به دست آمده از مقایسه شبکه های عصبی

کانولوشن با معماری متفاوت

(Table-3): The results obtained from the comparison of convolutional neural networks with different architectures

زمان- برای هر عکس	F1	RE	PR	معیار روش
۰/۰۷۶	۰/۵۷	۹۲/۹۹	۹۱/۳۵	روش پیشنهادی
۰/۰۸۸	۰/۶۱	۹۰/۳۰	۸۹/۹۵	روش مرجع [11]

در جدول (۴) مقایسه بین روش پیشنهادی و برخی روش های تشخیص تصاویر نامتعارف انجام شده است. با توجه به نتایج به دست آمده در جدول (۴)، روش یادگیری عمیق عملکرد مناسبی دارد و شبکه عصبی کانولوشن با معماری پیشنهادی، بهترین عملکرد را از خود نشان داده است. در فیلترینگ هوشمند تصاویر، علاوه بر تشخیص

صحيح تصاویر نامتعارف، تشخیص صحيح تصاویر متعارف بسیار حائز اهمیت است که روش پیشنهادی در هر دو قسمت، بهتر از همه روش های مطرح شده عمل می کند. در این مقاله، روش های انتخاب شده برای مقایسه در دسته های مختلف روش های شناسایی تصاویر نامتعارف (در کارهای پیشین معرفی شده) وجود دارند. به مانند مرجع [11] از شبکه عصبی کانولوشن با معماری AlexNet استفاده شده است که در این مقاله برای شناسایی تصاویر نامتعارف به کار برده شده است. از آنجایی که تاکنون پژوهشی برای شناسایی تصاویر نامتعارف توسط شبکه عصبی کانولوشن AlexNet صورت نگرفته است، پیاده سازی و نتایج این روش توسط نویسنده این مقاله انجام شده است.

شایان به ذکر است که نتایج جدول (۴)، از مرجع [42] برگرفته شده است. این نتایج بر روی پایگاه داده یکسان، حاصل شده است.

(جدول-۴): نتایج به دست آمده از روش های مختلف برای طبقه بندی تصاویر نامتعارف

(Table-4): Results obtained from different methods for categorizing unconventional images

مرجع [55]	مرجع [54]	مرجع [53]	مرجع [42]	مرجع [11]	روش پیشنهادی	روش معیار ارزیابی
۹۰/۹۳	۷۸/۳۷	۸۳/۶۵	۸۶/۹۷	۸۹/۵	۹۱/۳۵	TP
۱۸/۹۴	۱۶/۹۸	۶/۶۴	۵/۳۸	۹/۲۱	۵/۱۱	FP
۹/۰۷	۲۱/۶۳	۱۶/۳۵	۱۳/۰۳	۱۰/۵	۸/۶۵	FN
۸۱/۰۶	۸۳/۰۲	۹۳/۳۶	۹۴/۶۲	۹۰/۷۹	۹۴/۸۹	TN

## ۵- نتیجه‌گیری

امروزه با پیشرفت اینترنت و رسانه‌های تحت وب، توزیع و اشتراک منابع اطلاعاتی به مانند تصویر، در حال افزایش است. اشتراک منابع، خطرات و مشکلاتی را نظیر دسترسی افراد پایین‌تر از سن قانونی به تصاویر نامتعارف را ایجاد می‌کند. به همین منظور، در این مقاله به طراحی سامانه‌ای هوشمند برای تحلیل و فیلترینگ تصاویر تحت وب پرداخته شد که در حجم بالای داده‌ها نیز کارا است. در همین راستا، در این مقاله روش جدیدی بر پایه شبکه‌های عصبی کانولوشن با معماری نوین و مقاوم نسبت به حجم بالای داده، معرفی شد. نتایج به دست آمده نشان‌دهنده بهبود عملکرد روش پیشنهادی، نسبت به روش‌های دیگر بر روی مجموعه داده به نسبت بزرگ است.

## 6- References

## ۶- مراجع

- [1] Richmond, R., Facebook's new way to combat child pornography. New York Times, 2011
- [2] <http://www.dailyinfographic.com/the-stats-on-internet-pornography-nfographic>, accessed 2017/1/19.
- [3] Malamuth, N.M., T. Addison, and M. Koss, Pornography and sexual aggression: Are there reliable effects and can we understand them? Annual review of sex research, 2000. 11(1): p. 26-91.
- [4] Malamuth, N.M., Criminal and noncriminal sexual aggressors. Annals of the New York Academy of Sciences, 2003. 989(1): p. 33-58.
- [5] Alexy, E.M., A.W. Burgess, and R.A. Prentky, Pornography use as a risk marker for an aggressive pattern of behavior among sexually reactive children and adolescents. Journal of the American Psychiatric Nurses Association, 2009. 14(6): p. 442-453
- [6] Carter, D.L., R.A. Prentky, R.A. Knight, P.L. Vanderveer, and R.J. Boucher, Use of pornography in the criminal and developmental histories of sexual offenders. Journal of Interpersonal Violence, 1987. 2(2): p. 196-211.
- [7] Lin, Y.-C., H.-W. Tseng, and C.-S. Fuh. Pornography detection using support vector machine. in 16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003). 2003.
- [8] Zuo, H., W. Hu, and O. Wu. Patch-based skin color detection and its application to pornography image filtering. in Proceedings of the 19th

مرجع [53] روشی هوشمند بر پایه بافت و ویژگی‌های بصری معرفی کرده است. ویژگی‌های بصری استفاده شده از خصیصه‌هایی بر پایه شکل نواحی مرتبط به پوست بدن استخراج شده است. در نهایت تصاویر نامتعارف بر پایه ویژگی‌های معرفی شده توسط یک طبقه‌بند با ساختار سلسله‌مراتبی شناسایی شده‌اند. مرجع [54] نیز از ویژگی‌های شکل نواحی پوست افراد استفاده کرده است. این مقاله یک سامانه طبقه‌بند با چندین بیزا برای شناسایی نواحی پوست معرفی کرده است. مرجع [55] از روش‌های بازبایی تصاویر بر پایه محتوا استفاده کرده است. در این روش ابتدا پس‌زمینه حذف شده و نواحی مرتبط به پیسکل‌های پوست استخراج شده است؛ سپس تصویر نامتعارف توسط ویژگی‌های رنگ، بافت و ویژگی شکل پیشنهادی مقاله به کمک معیار شباهت تشخیص داده‌اند.

در مجموعه داده استفاده شده، چگالی توزیع تصاویر، متعادل است؛ اما در وب، توزیع تصاویر نامتعارف و متعارف یکسان نیستند و تعداد تصاویر متعارف بسیار بیشتر است؛ به همین دلیل مجموعه داده‌ای برای این منظور از اینترنت جمع‌آوری شده و نتایج یادگیری عمیق، با معماری پیشنهادی و معماری AlexNet بر روی این مجموعه داده نیز ارزیابی شده است. در مجموعه داده جدید از ۳۵۶۰۰ تصویر استفاده شده است که ۲۶۷۰۰ نمونه از آن‌ها، تصاویر متعارف و ۸۹۰۰ نمونه باقیمانده، تصاویر نامتعارف هستند. نتایج به دست آمده بر روی این مجموعه داده، در جدول (۵) نشان داده شده است.

با توجه به نتایج به دست آمده در این آزمایش، شبکه عصبی با معماری پیشنهادی، عملکرد بهتری بر روی مجموعه داده با چگالی توزیع نامتعادل، دارد.

(جدول ۵): نتایج به دست آمده از مقایسه شبکه‌های

عصبی کانولوشن با معماری متفاوت در مجموعه داده

جدید با توزیع نامتعادل

(Table-5): The results obtained from the comparison of convolutional neural networks with different architectures in the new data set with unbalanced distribution

F1	RE	PR	معیار ارزیابی روش
۸۹/۷۹	۸۷/۳۵	۹۲/۶۰	روش پیشنهادی
۸۳/۶۲	۷۹/۹۵	۸۷/۶۶	روش مرجع [11]

<sup>1</sup> Multi-Bayes

- pornography. in 2014 IEEE International Conference on Image Processing (ICIP). 2014. IEEE.
- [23] Ulges, A. and A. Stahl. Automatic detection of child pornography using color visual words. in 2011 IEEE International Conference on Multimedia and Expo. 2011. IEEE.
- [24] Sui, L., J. Zhang, L. Zhuo, and Y. Yang, Research on pornographic images recognition method based on visual words in a compressed domain. IET image processing, 2012. 6(1): p. 87-93.
- [25] Dong, K., L. Guo, and Q. Fu. An adult image detection algorithm based on Bag-of-Visual-Words and text information. in 2014 10th International Conference on Natural Computation (ICNC). 2014. IEEE.
- [26] Lowe, D.G., Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004. 60(2): p. 91-110.
- [27] Zheng, H., M. Daoudi, and B. Jedynek, Blocking adult images based on statistical skin detection. ELCVIA: electronic letters on computer vision and image analysis, 2004. 4(2): p. 001-14.
- [28] Wang, M., X.-S. Hua, J. Tang, and R. Hong, Beyond distance measurement: constructing neighborhood similarity for video annotation. IEEE Transactions on Multimedia, 2009. 11(3): p. 465-476.
- [29] Wang, M., X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, Unified video annotation via multigraph learning. IEEE Transactions on Circuits and Systems for Video Technology, 2009. 19(5): p. 733-746.
- [30] Wang, M., X.-S. Hua, T. Mei, R. Hong, G. Qi, Y. Song, and L.-R. Dai, Semi-supervised kernel density estimation for video annotation. Computer Vision and Image Understanding, 2009. 113(3): p. 384-396.
- [31] Ries, C.X. and R. Lienhart, A survey on visual adult image recognition. Multimedia tools and applications, 2014. 69(3): p. 661-688.
- [32] Duan, L., G. Cui, W. Gao, and H. Zhang. Adult image detection method base-on skin color model and support vector machine. in Asian conference on computer vision. 2002.
- [33] Yin, H., X. Huang, and Y. Wei, SVM-based pornographic images detection, in Software Engineering and Knowledge Engineering: Theory and Practice. 2012, Springer. p. 751-759.
- [34] Zaidan, A., H.A. Karim, N. Ahmad, B. Zaidan, and M.M. Kiah, Robust Pornography Classification Solving the Image Size Variation Problem Based on Multi-Agent Learning. Journal of Circuits, Systems and Computers, 2015. 24(02): p. 1550023.
- international conference on World wide web. 2010. ACM.
- [9] Largillier, T., G. Peyronnet, and S. Peyronnet, Efficient filtering of adult content using textual information. Murdock et al.[7], 2016: p. 14-17.
- [10] Rowley, H.A., Y. Jing, and S. Baluja. Large scale image-based adult-content filtering. in VISAPP (1). 2006. Citeseer.
- [11] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in Advances in neural information processing systems. 2012.
- [12] Forsyth, D.A. and M.M. Fleck. Identifying nude pictures. in Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on. 1996. IEEE.
- [13] Fleck, M.M., D.A. Forsyth, and C. Bregler. Finding naked people. in European Conference on Computer Vision. 1996. Springer.
- [14] Hu, W., O. Wu, Z. Chen, Z. Fu, and S. Maybank, Recognition of pornographic web pages by classifying texts and images. IEEE transactions on pattern analysis and machine intelligence, 2007. 29(6): p. 1019-1034.
- [15] Zhuo, L., J. Zhang, Y. Zhao, and S. Zhao, Compressed domain based pornographic imagerecognition using multi-cost sensitive decision trees. Signal Processing, 2013. 93(8): p. 2126-2139.
- [16] Wang, M. and X.-S. Hua, Active learning in multimedia annotation and retrieval: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 2011. 2(2): p. 10.
- [17] Li, F.-f., S.-w. Luo, X.-y. Liu, and B.-j. Zou, Bag-of-visual-words model for artificial pornographic images recognition. Journal of Central South University, 2016. 23(6): p. 1383-1389.
- [18] Baeza-Yates, R. and B. Ribeiro-Neto, Modern information retrieval. Vol. 463. 1999: ACM press New York.
- [19] Zhang, J., L. Sui, L. Zhuo, Z. Li, and Y. Yang, An approach of bag-of-words based on visual attention model for pornographic images recognition in compressed domain. Neurocomputing, 2013. 110: p. 145-152.
- [20] Wang, Y., L. Ning, and W. Gao, Detecting pornographic images with visual words. Transactions of Beijing Institute of Technology, 2008. 28(5): p. 410-13.
- [21] Gao, Y., M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu, Visual-textual joint relevance learning for tag-based social image search. IEEE Transactions on Image Processing, 2013. 22(1): p. 363-376.
- [22] Sae-Bae, N., X. Sun, H.T. Sencar, and N.D. Memon. Towards automatic detection of child

- Recognition, 2002. Proceedings. 16th International Conference on. 2002. IEEE.
- [48] Le Cun, B.B., J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. Handwritten digit recognition with a back-propagation network .in Advances in neural information processing systems. 1990. Citeseer.
- [49] LeCun, Y. and Y. Bengio, Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 1995. 3361(10): p. 1995.
- [50] Fei-Fei, L. ImageNet: crowdsourcing, benchmarking & other cool things. in CMU VASC Seminar. 2010.
- [51] Wang, J.Z., J. Li, G. Wiederhold, and O. Firschein, System for screening objectionable images. Computer Communications, 1998. 21(15): p. 1355-1360.
- [52] Platzer, C., M. Stuetz, and M. Lindorfer. Skin sheriff: a machine learning solution for detecting explicit images. in Proceedings of the 2nd international workshop on Security and forensics in communication systems. 2014. ACM.
- [53] Ahmadi, A., M. Fotouhi, and M. Khaleghi , Intelligent classification of web pages using contextual and visual features. Applied Soft Computing, 2011. 11(2): p. 1638-1647.
- [54] Zheng, Q.-F., W. Zeng, W.-Q. Wang, and W. Gao, Shape-based adult image detection. International Journal of Image and Graphics, 2006. 6(01): p. 115-124.
- [55] Shih, J.-L., C.-H. Lee, and C.-S. Yang, An adult image identification system employing image retrieval technique. Pattern Recognition Letters, 2007. 28(16): p. 2367-2374.
- [35] Wu, O., H. Zuo, W. Hu, M. Zhu, and S. Li. Recognizing and filtering web images based on people's existence. in Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01. 2008. IEEE Computer Society.
- [36] Li, D., N. Li, J. Wang, and T. Zhu, Pornographic images recognition based on spatial pyramid partition and multi-instance ensemble learning. Knowledge-Based Systems, 2015. 84: p. 214-223.
- [37] Zaidan, A.A., N.N. Ahmad, H.A. Karim, M. Larbani, B.B. Zaidan, and A. Sali, On the multi-agent learning neural and Bayesian methods in skin detector and pornography classifier: An automated anti-pornography system. Neurocomputing, 2014. 131: p. 397-418.
- [38] Yin, H., X. Xu, and L. Ye. Big skin regions detection for adult image identification. in Digital Media and Digital Content Management (DMDCM), 2011 Workshop on. 2011. IEEE.
- [39] Bosson, A., G.C. Cawley, Y. Chan, and R. Harvey. Non-retrieval: blocking pornographic images. in International Conference on Image and Video Retrieval. 2002. Springer.
- [40] Zhang, J., L. Sui, L. Zhuo, and Z. Li, Pornographic image region detection based on visual attention model in compressed domain. IET Image Processing, 2013. 7(4): p. 384-391.
- [41] Bozorgi, M., M.A. Maarof, and L.Z. Sam. Multi-classifier Scheme with Low-Level Visual Feature for Adult Image Classification. in International Conference on Software Engineering and Computer Systems. 2011. Springer.
- [42] Kia, S.M., H. Rahmani, R. Mortezaei, M.E. Moghaddam, and A. Namazi, A Novel Scheme for Intelligent Recognition of Pornographic Images. arXiv preprint arXiv:1402.5792, 2014.
- [43] Lienhart, R. and R. Hauke. Filtering adult image content with topic models. in 2009 IEEE International Conference on Multimedia and Expo. 2009. IEEE.
- [44] Islam, M., P. Watters, J. Yearwood, M. Hussain, and L.A. Swarna, Illicit Image Detection Using Erotic Pose Estimation Based on Kinematic Constraints, in Innovations and Advances in Computer, Information, Systems Sciences, and Engineering. 2013, Springer. p. 481-495.
- [45] Bengio, Y., Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2009. 2(1): p. 1-127.
- [46] Peter, Z.C., Building High-level Features Using Large Scale Unsupervised Learning.
- [47] Fasel, B. Robust face analysis using convolutional neural networks. in Pattern



**علی قنبری سرخی** مدرک کارشناسی خود را در رشته کامپیوتر گرایش نرم‌افزار از دانشگاه علم و صنعت ایران در سال ۱۳۸۹ دریافت کرد؛ سپس در سال ۱۳۹۱ موفق به اخذ مدرک کارشناسی ارشد در گرایش هوش مصنوعی از دانشگاه صنعتی شاهرود شد. وی در حال حاضر نیز، دانشجوی مقطع دکترا در دانشگاه صنعتی شاهرود در گرایش هوش مصنوعی است. موضوع پایان‌نامه دکترا ایشان، "دسته‌بندی محتوایی تصویر برپایه مدل‌های موضوعی" است. علایق پژوهشی او پردازش تصویر و یادگیری عمیق است. نشانی رایانامه ایشان عبارت است از:

Ali.ghanbari289@gmail.com



**منصور فاتح** مدرک کارشناسی خود را در رشته مهندسی برق از دانشگاه صنعتی شاهرود در سال ۱۳۸۶ دریافت کرد؛ سپس کارشناسی ارشد و دکتری خود را در رشته‌های مهندسی پزشکی و الکترونیک دیجیتال در سال‌های ۱۳۸۸ و ۱۳۹۳ از دانشگاه تربیت مدرس دریافت کرد. پروژه کارشناسی ارشد خود را با عنوان " بررسی نقش و اثر نور پلاریزه در درماتوسکپی از بدن با استفاده از شبیه‌سازی" و پروژه دکتری خود را با عنوان " خواندن خودکار نقشه‌های دستی فرش" انجام رسانید. از سال ۱۳۹۴ ایشان عضو هیئت علمی دانشگاه صنعتی شاهرود بوده و زمینه پژوهشی ایشان پردازش تصویر و یادگیری تقویتی است. نشانی رایانامه ایشان عبارت است از:

**mansoor fateh@shahroodut.ac.ir**



**حمید حسن پور** استاد تمام دانشکده مهندسی کامپیوتر دانشگاه شاهرود است. ایشان در سال ۱۳۷۲ مدرک کارشناسی مهندسی کامپیوتر خود را از دانشگاه علم و صنعت و در سال ۱۳۷۵ مدرک کارشناسی ارشد خود را در گرایش هوش ماشین از دانشگاه صنعتی امیرکبیر دریافت کرد. در سال ۱۳۸۳ موفق به اخذ مدرک دکترای خود از دانشگاه صنعتی کوئینزلند استرالیا در گرایش پردازش سیگنال شد. از سال ۱۳۸۴ الی ۱۳۸۶ نامبرده به‌عنوان عضو هیئت علمی در دانشکده مهندسی برق و کامپیوتر دانشگاه صنعتی بابل فعالیت داشت؛ سپس به دانشکده مهندسی کامپیوتر دانشگاه شاهرود انتقال یافت. زمینه‌های علمی مورد علاقه ایشان پردازش سیگنال، پردازش تصویر داده‌کاوی و پردازش متن است. نشانی رایانامه ایشان عبارت است از:

**h.hassanpour@shahroodut.ac.ir**