



Learning of the 5-Finger Robot Hand Using Deep Learning for Stable Grasping

ARTICLE INFO

Article Type

Original Research

Authors

Heidari H,^{1*}
Ghahri Saremi Tay,¹
Ghahri Saremi Tah.¹

How to cite this article

Heidari H, Ghahri Saremi Tay, Ghahri Saremi Tah, Learning of the 5-finger robot hand using deep learning for stable grasping, Modares Mechanical Engineering, 2023;23(05):307-316.

¹ Mechanical Engineering Department, Malayer University, Malayer, Iran

*Correspondence

Address: Faculty of Mechanical Engineering, Malayer University, Malayer, Iran
hr.heidari@malayeru.ac.ir

Article History

Received: January 03, 2023
Accepted: March 29, 2023
ePublished: May 20, 2023

ABSTRACT

The human hand is one of the most complex organs of the human body, capable of performing skilled tasks. Manipulation, especially grasping is a critical ability for robots. However, grasping objects by a robot hand is a challenging issue. Many researchers have used deep learning and computer vision methods to solve this problem. This paper presents a humanoid 5-degree-of-freedom robot hand. The robotic hand is made using a 3D printer and 5 servo motors are used to move the fingers. In order to simplify the robotic hand, a tendon-based transmission system was chosen that allows the robot's fingers to flexion and extension. The purpose of this article is to use deep learning algorithm to grasping different objects semi-automatically. In this regard, a convolutional neural network structure is trained with more than 600 images. These images were collected by a camera mounted on the robot's hand. Then, the performance of this algorithm is tested on different objects in similar conditions. Finally, the robot hand is able of successfully grasping with 85% accuracy.

Keywords Deep learning, Grasping, Convolutional neural network, Robot hand, Computer vision

CITATION LINKS

1- The anatomy and mechanics of the human hand. 2- State of the art in hand and finger modeling and animation. 3- A neural model of human prehension. 4- Effects of end-goal on hand shaping. 5- On grasp choice, grasp models, and the design of hands for manufacturing tasks. 6- Dexterous grippers: Putting nonholonomy to work for fine manipulation. 7- An introductory revision to humanoid robot hands. 8- ADA assembly instructions. 9- INMOOV robotic hand. 10- Review of gifu hand and its application. 11- DLR-Hand II: Next generation of a dextrous robot hand. 11- DLR-Hand II: Next generation of a dextrous robot hand. 12- Learning grasps with topographic features. 13- Deep learning for detecting robotic grasps. 14- 3D object recognition with deep belief nets. 15- Convolutional networks for images, speech, and time series. 16- A tutorial survey of architectures, algorithms, and applications for deep learning. 17- Imagenet classification with deep convolutional neural networks. 18- Neural networks and their applications. 19- Very deep convolutional networks for large-scale image recognition. 20- ImageNet Large Scale Visual Recognition Challenge. 21- An adaptive learning mechanism for teaching a robotic hand to grasp. 22- Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. 23- Design and evaluation of anthropomorphic robotic hand for object grasping and shape recognition. 24- A logical calculus of the ideas immanent in nervous activity. 25- Surveys and Presentations in Information Systems Engineering. 26- Deep convolution neural networks in computer vision: a review. 27- Towards better analysis of deep convolutional neural networks. 28- Mathworks: Introducing Deep Learning with MATLAB. 29- Huber loss. 30- Enabling grasping of unknown objects through a synergistic use of edge and surface information. 31- One-shot learning and generation of dextrous grasps for novel objects.

یادگیری دست ربات ۵ انگشتی با استفاده از یادگیری عمیق به منظور گرفتن پایدار

حمیدرضا حیدری^{۱*}، طاهره قهری صارمی^۱، طیبه قهری صارمی^۱

^۱ دانشکده فنی و مهندسی، دانشگاه ملایر، ملایر، ایران

چکیده

دست انسان یکی از پیچیده‌ترین اندام‌های بدن انسان است که قادر به انجام وظایف ماهرانه می‌باشد. ماهرانه عمل کردن به ویژه گرفتن یک توانایی حیاتی برای ربات‌ها محسوب می‌شود. با این حال، گرفتن اشیاء توسط دست ربات یک مسئله چالش برانگیز است. بسیاری از محققان از روش‌های یادگیری عمیق و بینایی کامپیوتری برای حل این مسئله استفاده کرده‌اند. این مقاله یک دست ربات ۵ درجه آزادی انسان‌نما را ارائه می‌دهد. دست رباتیک با استفاده از پربینتر سه بعدی ساخته شده و برای حرکت انگشتان از ۵ سرو موتور استفاده می‌شود. به منظور سادگی دست رباتیک، سیستم انتقال مبتنی بر تاندون انتخاب شده که به انگشتان دست ربات اجازه خمش و کشش می‌دهد. هدف این مقاله استفاده از الگوریتم یادگیری عمیق برای گرفتن نیمه خودکار اشیاء مختلف می‌باشد. در این راستا یک ساختار شبکه عصبی کانولوشن با بیش از ۶۰۰ تصویر آموزش داده می‌شود. این تصاویر توسط یک دوربین نصب شده بر روی دست ربات جمع آوری شده است. سپس عملکرد این الگوریتم در شرایط مشابه روی اشیاء مختلف آزمایش می‌شود. در نهایت دست رباتیک قادر به گرفتن موفقیت آمیز با دقت ۸۵ درصد می‌باشد.

کلیدواژه‌ها: یادگیری عمیق، گرفتن، شبکه عصبی کانولوشن، دست ربات، بینایی کامپیوتر

تاریخ دریافت: ۱۴۰۱/۱۱/۱۳

تاریخ پذیرش: ۱۴۰۲/۰۱/۰۹

*نویسنده مسئول: hr.heidari@malayeru.ac.ir

۱- مقدمه

دست انسان از مهارت بالایی برخوردار است. محققان به دنبال تکرار حرکات در ساختارهای مکانیکی دست انسان به نام دست انسان‌نما یا دست رباتیک هستند. مطالعه دست‌های رباتیک یک زمینه وسیع و چندوجهی است که از مکانیک و ریاضیات تا تئوری کنترل و یادگیری ماشین را در بر می‌گیرد. کاربرد دست رباتیک را می‌توان در جست و جو، تعامل با محیط و انسان‌ها و ماهرانه عمل کردن (Manipulation) خلاصه نمود. با توجه به گستردگی زمینه، هنگام بحث درباره‌ی دست‌های رباتیک، این بررسی فقط بر روی دست‌های ربات با قابلیت ماهرانه عمل کردن متمرکز می‌شود. ماهرانه عمل کردن یک توانایی حیاتی برای ربات‌ها محسوب می‌شود. این توانایی بر روی ربات‌ها، آن‌ها را قادر می‌سازد تا به انسان‌ها کمک کنند و برای انجام وظایف پیچیده‌تر در زندگی روزمره و صنعت جایگزین شوند. نمونه‌ای از ماهرانه عمل کردن ربات، عمل گرفتن و رها کردن اشیاء می‌باشد. عمل گرفتن دست ربات یک کار بسیار پیچیده است و موفقیت آن به عوامل متعددی مانند شکل و موقعیت یک شیء بستگی دارد. در این راستا

دست‌های ماهر در کاربردهای مختلف گرفتن بررسی شده‌اند. با افزایش قابل توجه قدرت محاسباتی نسبت به سخت افزار، هوش مصنوعی که توسط روش‌های یادگیری عمیق (Deep learning) ارائه می‌شود، به سرعت در زمینه‌ی گرفتن دست ربات نفوذ پیدا کرده است. هوش مصنوعی شاخه‌ای از علوم کامپیوتر است که هدف آن شبیه سازی هوش انسانی در ماشین‌ها می‌باشد. از کاربردهای دست ربات نیز می‌توان به جراحی، پروتز و صنعت تولید اشاره نمود. افزودن یک ماژول بینایی با الهام از جدیدترین راه حل‌های یادگیری عمیق به بهبود پروتزهای فعلی کمک می‌کند.

آناتومی دست بسیار پیچیده و جذاب است. دست انسان از ۸ استخوان برای مچ دست، ۵ استخوان در کف دست و ۱۴ استخوان برای انگشتان تشکیل شده است^[1]. مطابق شکل ۱ می‌توان مفاصل انگشتان دست را در طبقات زیر دسته بندی نمود: مفاصل متاکارپو فالانژیال (Metacarpo-phalangeal)، مفاصل اینترفالانژیال پروگزیمال (Proximal intra-phalangeal) و مفاصل اینترفالانژیال دیستال (Distal intra-phalangeal). انگشت اشاره، وسط، حلقه و کوچک هر کدام دارای ۴ درجه آزادی هستند که شامل ۱ درجه آزادی (خمش/کشش) در مفصل اینترفالانژیال دیستال، ۱ درجه آزادی (خمش/کشش) در مفصل اینترفالانژیال پروگزیمال و ۲ درجه آزادی (اداکشن (Adduction) / اداکشن (Abduction) و خمش/کشش) در مفصل متاکارپو فالانژیال می‌باشد. انگشت شست با ۲ مفصل دارای ۵ درجه آزادی است که علاوه بر ۱ درجه آزادی (خمش/کشش) در مفصل اینترفالانژیال دیستال و ۲ درجه آزادی (اداکشن / اداکشن شعاعی و خمش/کشش) در مفصل متاکارپو فالانژیال، دارای ۱ درجه آزادی برای حرکات چرخشی و ۱ درجه آزادی برای اداکشن/اداکشن کف دست می‌باشد. حرکت انگشت شست نسبت به انگشت اشاره به عنوان اداکشن / اداکشن شعاعی و حرکت انگشت شست نسبت به کف دست به عنوان اداکشن/اداکشن کف دست شناخته می‌شود. حرکات انتقالی و چرخشی کف دست نیز در مجموع با ۶ درجه آزادی مشخص می‌شود. بنابراین به طور کلی، دست انسان دارای ۲۷ درجه آزادی است^[2].

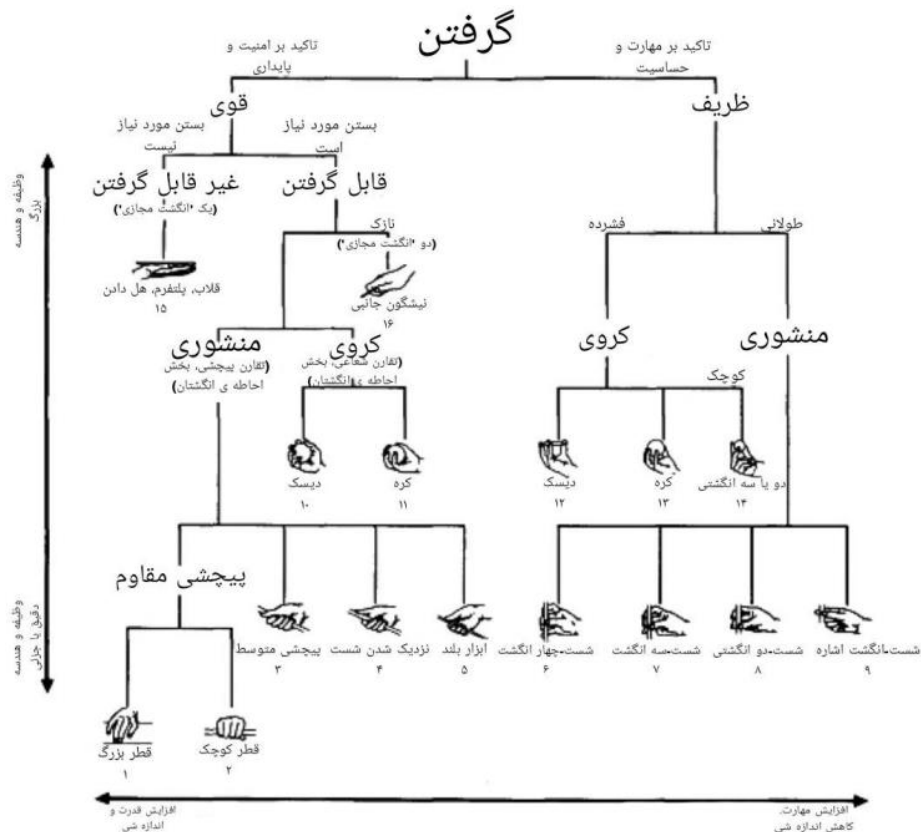


شکل ۱) شماتیک مفاصل دست انسان^[2]

شکل شیء تطبیق می‌یابد. گریپ‌های سه انگشتی موازی دارای انگشتانی هستند که به سمت نقطه میانی مشترک حرکت می‌کنند و به طور خودکار متمرکز می‌شوند. با این حال شبیه گریپ‌های دو انگشتی موازی برای گرفتن طیف وسیعی از اشیاء نامناسب هستند. گریپ‌های سه انگشتی تطبیقی انعطاف پذیری و قابلیت ایمنی بیشتری را ایجاد می‌کنند. دست رباتیک شبیه دست انسان نه تنها به دلیل زیبایی، بلکه توسط سال‌ها تکامل برای بهره برداری از طراحی طبیعی به ابزاری همه کاره تبدیل شده است و در بسیاری از کارهای روزانه کمک کننده می‌باشد [7]. در دست رباتیک ۵ Ada درجه آزادی از محرک‌های خطی و تاندون‌ها به عنوان انتقال حرکت و از سنسورهای عضلانی برای کنترل استفاده می‌شود [8]. InMoov یک دست ربات منبع باز است که با استفاده از پرینتر سه بعدی ساخته می‌شود. در دست ربات InMoov از ۵ سرو موتور استفاده شده که به عنوان محرک عمل می‌کنند و با استفاده از میکروکنترلر آردوینو کنترل می‌شوند [9]. دست ربات Gifu شبیه یک دست انسانی طراحی شده است. سنسورهای نیرو در نوک انگشتان و سنسورهای لمسی توزیع شده در سراسر انگشتان و کف دست این ربات به گرفتن اشیاء کمک کرده است [10]. دست ربات DLR-II با ۱۳ درجه آزادی از سه انگشت و یک شست تشکیل شده است. موتورهای DC به عنوان محرک عمل می‌کنند و انتقال حرکت در مفاصل توسط چرخ دنده‌ها انجام می‌شود که امکان کشش و خمش انگشتان را فراهم می‌کند [11].

این ساختار پیچیده به دست انسان، توانایی حرکت ماهرانه می‌دهد. حرکات مختلف دست برای عمل گرفتن ضروری می‌باشد. عمل گرفتن به عنوان "ایمن نگه داشتن یک شیء در حالت ثابت دست" تعریف شده است [3]. مطالعات نشان داده است که وضعیت گرفتن به ویژگی‌های فیزیکی شیء به عنوان مثال شکل و اندازه بستگی دارد [4]. در این راستا کاتکاسکی (Cutkosky) [5] طبقه بندی بسیار جامع و دقیقی از عمل گرفتن دست انسان ارائه داد. او گرفتن را به دو دسته قوی و ظریف تقسیم بندی نمود که این دو دسته به چندین زیرمجموعه تقسیم می‌شود. گرفتن قوی به عنوان تماس کامل دست انسان با یک شیء تعریف می‌شود. در گرفتن ظریف، انگشت اشاره و شست بیشتر مورد استفاده قرار می‌گیرد. طبقه بندی کاتکاسکی در شکل ۲ آورده شده است.

درک مکانیزهای عمل گرفتن انسان برای انتقال این مهارت‌ها به دست‌های رباتیک، مورد توجه محققان قرار گرفته است [3]. برای طراحی دست‌های رباتیک ماهر دو نظریه فکری وجود دارد: اول ساخت دستی که شبیه دست انسان است و می‌تواند طیف وسیعی از کاربردها را داشته باشد و دوم طراحی یک گریپر که برای وظایف معین ایجاد شده است [6]. گریپ‌های دو و سه انگشتی به دو نوع موازی و تطبیقی تقسیم بندی شده‌اند. گریپ‌های دو انگشتی موازی به صورت موازی با یکدیگر حرکت می‌کنند و با طراحی نسبتاً ساده در اندازه و مواد مختلف، توانستند عمل گرفتن را انجام دهند. با این حال برای کارهایی که نیاز به گرفتن اشیاء مختلف دارند مناسب نمی‌باشند، در حالی که یک گریپر دو انگشتی تطبیقی با



شکل ۲) طبقه بندی گرفتن [5]

دانشگاه آکسفورد پیشنهاد شد. این مدل در چالش ILSVRC 2014 مقام دوم را کسب نمود که قادر به تشخیص ۱۰۰۰ شیء مختلف می‌باشد. گومز و همکاران [21] یک شبکه عصبی را پیشنهاد کردند که با قابلیت‌های یک دست رباتیک سازگار است و از اطلاعات حسی مختلف به عنوان ورودی استفاده می‌شود. لوین و همکاران [22] از ورودی بصری استفاده کردند. بر اساس اطلاعات بصری، احتمال گرفتن موفق پیش‌بینی می‌شود. آن‌ها به منظور آموزش شبکه عصبی کانولوشن از مجموعه داده‌ای با ۸۰۰۰۰۰ تلاش استفاده کردند. دواراجا و همکارانش [23] یک دست رباتیک چند انگشتی برای گرفتن اشیاء ایجاد کردند و از یک طبقه بندی کننده ماشین بردار پشتیبان (Support vector machine) برای تشخیص شکل شیء استفاده کردند. آن‌ها دست رباتیک طراحی شده را روی چند شیء آموزش دادند و به میانگین دقت تشخیص شکل شیء ۹۴/۴ درصد دست یافتند. هدف اصلی اکثر پروژه‌های تحقیقاتی پیشرفته‌ی جهان، دستیابی به عملکرد مستقل انگشتان و دقت در نگه داشتن اشیاء می‌باشد.

در این راستا به منظور گرفتن طیف وسیعی از اشیاء و زیبایی در طراحی، یک دست ربات ۵ انگشتی انتخاب شد. ساخت دست ربات با استفاده از روش پرینتر سه بعدی با هدف کاهش هزینه انجام می‌شود. سپس از شبکه‌های عصبی کانولوشن به منظور یادگیری گرفتن پایدار دست رباتیک استفاده خواهد شد تا باز و بسته شدن خودکار انگشتان تعیین شود. این مقاله چندین مزیت نسبت به روش‌های فعلی دارد. استفاده از یادگیری عمیق به دست ربات پیشنهادی منجر به اجتناب از کنترل انسانی شده و در عوض با توانایی بخشیدن به دست ربات، بهترین استراتژی گرفتن انجام می‌پذیرد. هم چنین بسیاری از روش‌های یادگیری عمیق از نظر محاسباتی سنگین هستند و به مقدار زیادی داده‌ی آموزشی نیاز دارند. در این مقاله با استفاده از ساختار شبکه عصبی از پیش آموزش دیده و تکنیک تنظیم دقیق، فقط لایه‌ی آخر آموزش داده می‌شود. استفاده از این رویکرد دیگر نیازی به آموزش شبکه با تعداد زیادی داده ندارد.

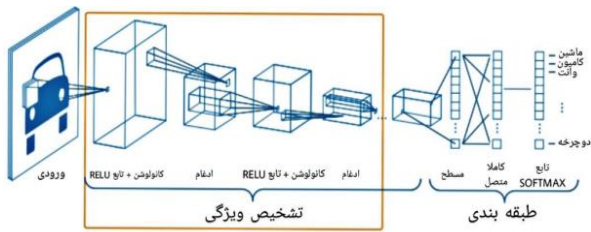
۲- شبکه‌های عصبی برای وظایف پیش بینی در گرفتن

در این بخش توضیح مختصری از برخی مفاهیم شبکه‌های عصبی و شبکه‌های عصبی کانولوشن ارائه می‌شود.

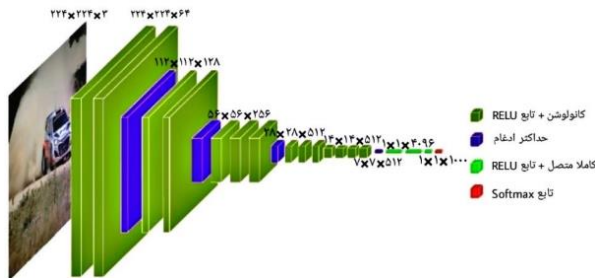
۲-۱- شبکه‌های عصبی

شبکه‌های عصبی اولین بار توسط وارن مک‌کالو و والتر پیترز در سال ۱۹۴۴ [24] ارائه شد. شبکه عصبی مصنوعی (Artificial Neural Networks (ANN)) یک الگوی پردازش اطلاعات می‌باشد که از سیستم‌های عصبی بیولوژیکی مانند مغز انسان الهام گرفته شده است. یک شبکه عصبی از تعداد زیادی عناصر پردازشی به هم پیوسته به نام نورون (Neurone) تشکیل شده است که به طور هماهنگ برای حل مسائل خاصی کار می‌کنند. شبکه عصبی در یک

در این راستا محققان ایده‌هایی برای گرفتن ماهرانه از دیدگاه فیزیک، هندسه و مدل‌سازی سینماتیک ارائه کردند [12]. با این حال الگوریتم‌های مبتنی بر فیزیک و هندسه به راه حل‌های سخت افزاری وابسته هستند و تعمیم این روش‌ها سخت می‌باشد. با افزایش قابل توجه قدرت محاسباتی نسبت به سخت افزار، هوش مصنوعی به سرعت در زمینه‌های مختلف تحقیقاتی نفوذ پیدا کرد. پیشرفت الگوریتم‌های مبتنی بر یادگیری عمیق، هوش مصنوعی را به سطح بالاتری رسانده است. به طور خاص گرفتن مبتنی بر شبکه عصبی (Convolutional neural network) کانولوشن با تصاویر RGB، موفقیت بالایی در بسیاری از کارها کسب نموده است [13]. تصاویر شبکه‌های عصبی عمیق معمولاً ماتریس‌هایی در مقیاس بزرگ هستند. با توجه به ابعاد بالای داده‌ها در تصاویر و نیاز به ظرفیت بالا برای یادگیری ویژگی‌های مختلف، معمولاً بیش از یک لایه پنهان در یک شبکه عصبی مورد نیاز است که منجر به مفهوم شبکه‌های عصبی عمیق می‌شود. با داشتن لایه‌های بیشتر، ویژگی‌های پیچیده‌تری به دست می‌آیند. هرچه ویژگی‌ها پیچیده‌تر باشند، پیش بینی‌های دقیق‌تر و وظایف پیچیده‌تری انجام می‌شود. شبکه‌های عمیق محاسبه بر هستند و راه حل‌های مختلفی برای رفع این مشکل پیشنهاد شده است. رمزگذار عمیق خودکار (Deep auto-encoders) [14] و شبکه‌های عصبی کانولوشن [15] از جمله راه حل‌های محبوب یادگیری عمیق برای تشخیص اشیاء هستند [16]. در مقایسه با سایر تکنیک‌های یادگیری عمیق، شبکه‌های عصبی کانولوشنی عمیق با توانایی بالا در یادگیری به خوبی برای مسائل طبقه‌بندی تصویر مناسب هستند. این شبکه‌ها عملکرد قابل توجهی را در انواع وظایف چالش برانگیز تشخیص اشیاء نشان می‌دهند [17]. در ساختار شبکه عصبی کانولوشن از قشر بینایی مغز الهام گرفته شده است. شبکه‌های عصبی معمولاً برای مسائل بینایی کامپیوتر (Computer vision) استفاده می‌شوند [18]. در سال ۲۰۰۹، مجموعه داده ImageNet در راستای پیشبرد تحقیقات یادگیری عمیق با مسئله تشخیص جمع‌آوری شد. ImageNet از هزاران تصویر در ۱۰۰۰ دسته تشکیل شده و در دسترس عموم قرار گرفته است. در سال ۲۰۱۲ معماری شبکه نوآورانه‌ای به نام AlexNet، مدل الهام گرفته از شبکه‌های عصبی کانولوشن در چالش ImageNet برنده شد. از آن زمان، شبکه‌های عصبی کانولوشن به استاندارد طلایی برای طبقه بندی تصاویر تبدیل شده‌اند. در ادامه‌ی موفقیت بزرگ AlexNet، انواع مختلفی از معماری‌های شبکه طراحی شد و محبوبیت یافت [17]. VGG (Visual Geometry Group) [19] و ResNet از جمله شبکه‌های عمیق معروف با کارایی بالا هستند. هر معماری طراحی خاصی را به همراه دارد که فرآیند یادگیری در شبکه را تسهیل می‌کند و عملکرد کلی تشخیص شیء را در ImageNet افزایش می‌دهد. به عنوان مثال، معماری VGG [20] که یکی از معماری‌های شبکه عصبی کانولوشن می‌باشد در سال ۲۰۱۴ توسط گروه هندسه بصری



شکل ۴) نمونه‌ای از ساختار شبکه عصبی کانولوشن [28]



شکل ۵) معماری VGG-16 [20]

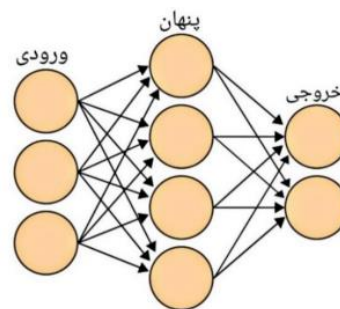
۳- طراحی و مدل‌سازی سیستم

۳-۱- ساخت دست ربات

برای این مطالعه، دست رباتیک منبع باز [9] ۵ انگشتی با هدف نمونه‌سازی سریع و هزینه پایین توسط پرینتر سه بعدی ساخته می‌شود. دست ربات دارای ۵ درجه آزادی می‌باشد. تمامی انگشتان به جز انگشت شست دارای طراحی و ابعاد یکسانی هستند که امکان ساخت آسان‌تر و قابلیت تعویض قطعات را فراهم می‌کند. ابعاد کلی انگشتان ۱/۵ سانتی متر (عرض) در ۹ سانتی متر (ارتفاع) و ابعاد انگشت شست ۱/۵ سانتی متر (عرض) در ۶/۵ سانتی متر (ارتفاع) می‌باشد. همانطور که در شکل ۶ مشاهده می‌شود، طرح ۴ انگشت پیشنهادی از سه مفصل شامل متاکارپو فالانژیال، اینترفالانژیال پروگزیمال، اینترفالانژیال دیستال و انگشت شست از دو مفصل شامل متاکارپو فالانژیال و اینترفالانژیال دیستال تشکیل شده است. تمامی انگشتان دست ربات قادر به حرکت خمش/کشش هستند. نداشتن حرکت ابداکشن/اداکشن کف دست برای انگشت شست، توانایی این انگشت را در مقایسه با انگشت شست انسان محدود می‌کند. در نتیجه انگشت شست رباتیک فقط با انگشت اشاره و یا انگشت وسط، عمل گرفتن را انجام می‌دهد. با این حال انگشت شست در ارتباط با انگشت اشاره، وسط یا هر دو انگشت، می‌تواند گرفتن موفقی را برای اشیای مختلف انجام دهد. در برخی حالات نیز انگشت شست باعث بهبود عمل گرفتن می‌شود.

نمونه دست ربات پیشنهادی شبیه دست انسان می‌باشد. در این راستا یک سیستم انتقال مبتنی بر تاندون انتخاب شد تا به انگشتان دست اجازه خمش و کشش دهد. برای این مکانیزم، از نخ ماهیگیری استفاده می‌شود. با افزودن یک پوشش سیلیکونی نرم به نوک انگشتان و یک ورقه فوم بر روی کف دست، چسبندگی و مهارت بیشتری فراهم می‌شود. برای حرکت انگشتان دست ربات

سری لایه طبقه بندی شده به عنوان ورودی، پنهان و خروجی مرتب شده‌اند. وظیفه لایه ورودی دریافت اطلاعات و پردازش آنهاست. وظیفه اصلی لایه‌های پنهان، تبدیل ورودی‌ها به چیزی است که لایه‌های خروجی بتوانند از آن استفاده کنند. در نهایت، لایه‌ی خروجی حاوی سیگنال‌هایی است که نشان می‌دهد چگونه شبکه به اطلاعاتی که یاد گرفته است پاسخ می‌دهد. روند رو به افزایشی برای استفاده از شبکه‌های عصبی در زمینه‌های مختلف، مانند طبقه‌بندی، کنترل و رباتیک وجود دارد [25]. شکل ۳ نمونه‌ای از یک شبکه عصبی را نشان می‌دهد که هر گره نشان دهنده یک نورون و هر فلش نشان دهنده اتصال از خروجی یک نورون به ورودی نورون دیگر است.





شکل ۳) ساختار شبکه عصبی

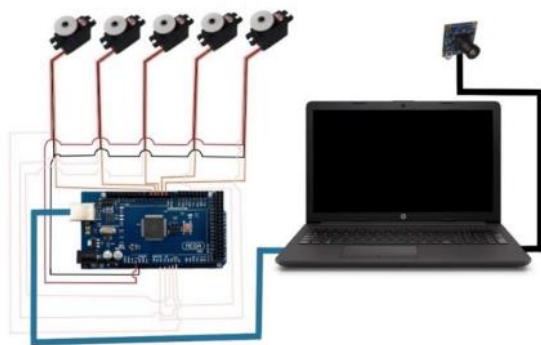
۲-۲- شبکه‌های عصبی کانولوشن

شبکه‌ی عصبی کانولوشن به روشی پیشرفته برای تشخیص اشیاء در جامعه بینایی کامپیوتر تبدیل شده است [26]. آموزش شبکه‌های عصبی کانولوشن نسبت به سایر شبکه‌های عصبی آسان‌تر است و پارامترهای کمتری برای تخمین دارند. این شبکه‌ها با گرفتن یک تصویر به عنوان ورودی، عملیات متوالی بر روی تصویر اعمال می‌کنند. شبکه عصبی کانولوشن شامل لایه‌های ورودی، کانولوشن (Convolutional)، ادغام (Pooling) و کاملاً متصل (Fully connected) می‌باشد. مراحل کانولوشن و ادغام را می‌توان چندین بار تکرار کرد و ویژگی‌ها را از یک تصویر ورودی استخراج نمود. در لایه کاملاً متصل از ویژگی‌های استخراج شده برای طبقه بندی استفاده می‌شود [27]. هرچه شبکه عمیق‌تر باشد، می‌تواند مسائل پیچیده‌تری را با دقت بالایی حل کند اما به زمان بیشتری برای پردازش نیاز دارد. شکل ۴ تصویری از ساختار یک شبکه عصبی کانولوشن را نشان می‌دهد. شبکه عصبی کانولوشن می‌تواند معماری‌های مختلفی داشته باشد. معماری VGG-16 با داشتن ۱۶ لایه به عنوان یک معماری استاندارد برای شبکه‌های عصبی کانولوشن به منظور تشخیص اشیاء تبدیل شده است. شکل ۵ معماری VGG-16 را نشان می‌دهد.

جدول ۱) اجزای مورد نیاز دست ربات

تعداد	تصویر قطعه	نام قطعه
۵		سروو موتور HS-81 Hitec
۱		ماژول دوربین ELP-USB100W05MT
۱		آرduino مگا ۲۵۶۰

وجود دارد. برد آرduino و دوربین از طریق USB به کامپیوتر متصل می‌شوند. شکل ۹ اتصالات سخت افزاری و ارتباط بین سروو موتور، آرduino، دوربین و کامپیوتر را نشان می‌دهد.

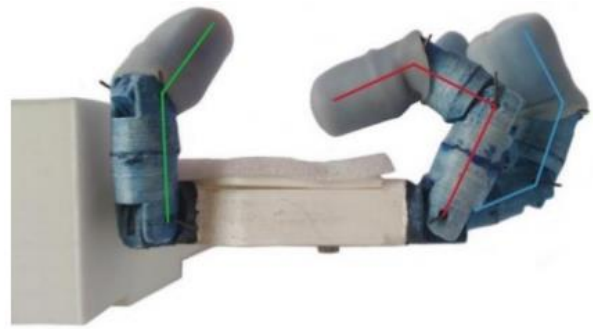


شکل ۹) اتصالات سخت افزاری

۴- روش تحقیق

۴-۱- جمع آوری داده‌های آموزشی

دو مجموعه داده شامل داده‌های ImageNet و داده‌های جمع آوری شده با دوربین به منظور آموزش دست ربات استفاده شد. داده‌های ImageNet یک پایگاه داده تصویری بزرگ است که در حال حاضر میلیون‌ها تصویر را شامل می‌شود. این مجموعه داده به طور گسترده ای برای آموزش شبکه‌های از پیش آموزش دیده مورد استفاده قرار می‌گیرد. داده‌های جمع آوری شده با دوربین شامل ۶۶۹ تصویر می‌باشد. این داده‌ها با استفاده از دوربین سوار شده بر



شکل ۶) شماتیکی از مفاصل دست ربات

از ۵ سروو موتور آنالوگ استفاده می‌شود. این سروو موتورها بر روی ساعد دست متصل می‌شوند. شکل ۷ نمای کلی دست ربات را نشان می‌دهد.



شکل ۷) نمای کلی دست ربات

به منظور جمع آوری اطلاعات بصری از یک دوربین استفاده می‌شود. این دوربین ۵ مگا پیکسلی دارای وضوح تصویر ۱۲۸۰x۷۲۰ است و با استفاده از یک قطعه چوب در فاصله‌ی ۴۰ سانتی متری به دست ربات متصل می‌شود. دلیل قرار گرفتن دوربین در فاصله‌ی ۴۰ سانتی متری، رزولوشن و پهنای دید دوربین می‌باشد تا اشیاء را واضح و کامل نشان دهد. تصویری از اتصال دوربین به دست ربات در شکل ۸ آورده شده است.

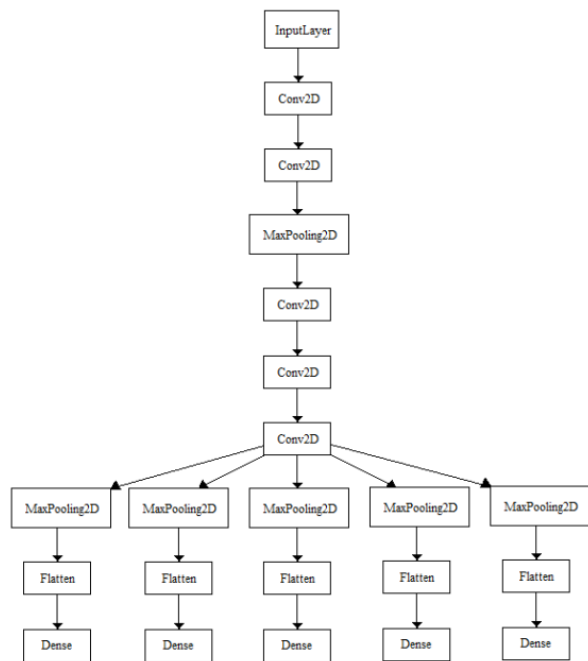


شکل ۸) اتصال دوربین به دست ربات

۳-۲- طراحی الکترونیکی

۵ سروو موتور که انگشتان دست را به حرکت در می‌آورند از طریق پین‌های PWM آرduino کنترل می‌شوند. اجزای مورد نیاز دست ربات در جدول ۱ آورده شده است. مقادیر موقعیت سروو موتورها از طریق پین‌های آنالوگ آرduino خوانده می‌شود. برد آرduino اطلاعات پردازش شده را از طریق پورت سریال دریافت می‌کند. از آنجایی که این برد قادر به تأمین برق ۵ سروو موتور به طور همزمان نیست، یک منبع تغذیه خارجی نیز

از لایه‌ی حداکثر ادغام، ۲ لایه‌ی کانولوشن با ۱۲۸ کانال و اندازه فیلتر ۳*۳ وجود دارد. مانند لایه‌ی قبل ۱ لایه‌ی حداکثر ادغام قرار می‌گیرد. پس از آن، ۳ لایه‌ی کانولوشن برای استخراج ویژگی و ۱ لایه‌ی حداکثر ادغام با ۲۵۶ کانال وجود دارد. در ادامه ۲ مجموعه از ۳ لایه‌ی کانولوشن و ۱ لایه‌ی حداکثر ادغام به منظور جلوگیری از بیش برآزش قرار دارند که هر کدام دارای ۵۱۲ کانال و اندازه فیلتر ۳*۳ می‌باشند. خروجی لایه‌ی کانولوشن در انتها (فیچرماپ) ۵۱۲، ۷، ۷ خواهد بود. لایه‌ی آخر که لایه‌ی کاملاً متصل است توسط داده‌های جمع‌آوری شده با دوربین دوباره آموزش داده می‌شود. در این راستا ۱ لایه‌ی مسطح (Flatten) اضافه می‌شود که بین لایه‌های کانولوشن و کاملاً متصل ارتباط برقرار می‌کند و به عنوان ورودی لایه‌ی کاملاً متصل در نظر گرفته می‌شود. لایه‌ی کاملاً متصل که شامل لایه‌های متراکم (Dense) می‌باشد، برای پیش‌بینی شبکه استفاده می‌شود. آخرین لایه‌ی کاملاً متصل به عنوان لایه‌ی خروجی عمل می‌کند.



شکل ۱۱) نمودار شبکه عصبی

۳-۴- آموزش شبکه

وابستگی به داده در آموزش، یکی از چالش برانگیزترین مسائل در یادگیری عمیق است به‌ویژه جمع‌آوری داده‌هایی که هزینه بر یا طولانی خواهد بود. این موضوع می‌تواند توسعه‌ی بسیاری از برنامه‌های کاربردی یادگیری عمیق را محدود کند. تحت این شرایط، یادگیری انتقالی می‌تواند سودمند باشد. تنظیم دقیق روشی برای به‌کارگیری یادگیری انتقالی با استفاده از شبکه‌ای است که قبلاً برای برخی از کارها آموزش دیده است و سپس معماری این شبکه بر روی کار دیگری تنظیم می‌شود. برای دستیابی به دقت بالاتر، در عوض آموزش یک مدل از ابتدا، یک

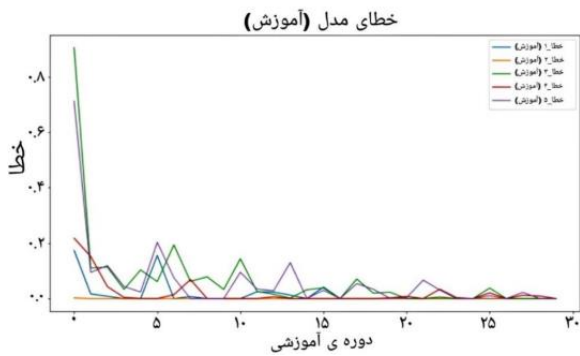
روی دست ربات جمع‌آوری می‌شود. در این راستا از یک چسب ماتیک با شکل استوانه‌ای در شرایط نوری و مکانی یکسان استفاده شد. شکل ۱۰ یک فریم از این مجموعه داده را نشان می‌دهد. در ادامه یک تصویر از دست در موقعیت باز و یک شیء در مجاورت آن ثبت می‌شود. محل قرار گرفتن چسب ماتیک در بالای کف دست ربات با زوایای مختلف و جابه‌جایی طولی و عرضی صورت می‌گیرد. سپس بستن دست با حرکت سروو موتورها صورت می‌گیرد و موقعیت هر انگشت به عنوان برچسب در آموزش شبکه استفاده می‌شود. برچسب‌ها به دو مرحله ۰ و ۱ تقسیم می‌شوند که به معنای دو کلاس "استفاده شده" و "استفاده نشده" است. داده‌های جمع‌آوری شده شامل یک فریم در هر تلاش برای گرفتن و یک فایل متنی است که مسیرهای کامل تمام تصاویر و برچسب‌های مربوط به آن‌ها را فهرست می‌کند.



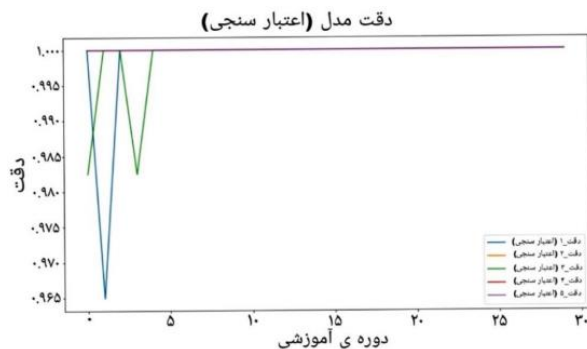
شکل ۱۰) نمونه‌ی یک فریم از مجموعه داده

۴-۲- معماری شبکه

معماری VGG-16 که از قبل بر روی مجموعه داده ImageNet آموزش داده شده است، برای انتخاب خودکار گرفتن استفاده می‌شود. انتخاب VGG-16 به دلیل معماری ساده و افزایش دقت نسبت به سایر معماری‌ها می‌باشد. ورودی شبکه تصویری VGG، تانسور (۳، ۲۲۴، ۲۲۴) می‌باشد که شامل تصاویر با اندازه ثابت ۲۲۴ در ۲۲۴ پیکسل و کانال‌های RGB است. ساختار VGG شامل یک سری از لایه‌های محاسباتی کانولوشن است که بعد از آن لایه‌های ادغام وجود دارند. آرایش لایه‌های کانولوشن و حداکثر ادغام به طور مداوم در کل معماری تکرار می‌شود. در پایان، لایه‌های کاملاً متصل برای خروجی در نظر گرفته می‌شود. نمودار شبکه عصبی استفاده شده در مقاله به صورت خلاصه در شکل ۱۱ آورده شده است. همانطور که در شکل ۱۱ مشاهده می‌شود بعد از لایه‌ی ورودی، تصاویر از ۲ لایه‌ی کانولوشن با ۶۴ کانال در اندازه فیلتر ۳*۳ عبور می‌کنند و به دنبال آن تابع فعال ساز ReLU انجام می‌شود که امکان آموزش سریع‌تر را فراهم می‌کند. سپس به منظور کاهش پیچیدگی، ۱ لایه‌ی حداکثر ادغام قرار می‌گیرد. پس



شکل ۱۳) نمودار خطای شبکه برای هر انگشت (۱: انگشت شست، ۲: انگشت اشاره، ۳: انگشت وسط، ۴: انگشت حلقه و ۵ انگشت کوچک)



شکل ۱۴) نمودار دقت شبکه برای هر انگشت (۱: انگشت شست، ۲: انگشت اشاره، ۳: انگشت وسط، ۴: انگشت حلقه و ۵ انگشت کوچک)

همانطور که مشاهده می‌شود دقت و خطای هر انگشت با ۳۰ دوره‌ی آموزشی قابل قبول می‌باشد. تابع خطا، پیش بینی را بر اساس اینکه چقدر با مقدار واقعی تفاوت دارد، در نظر می‌گیرد. دقت شبکه، تعداد پیش بینی‌هایی است که در آن مقدار پیش بینی شده با مقدار واقعی برابر است. با توجه به شکل ۱۳، در طول ۳۰ دوره‌ی آموزشی خطا به سرعت کاهش می‌یابد که نشان می‌دهد شبکه، تناسب خوبی با داده‌های آموزشی پیدا کرده است. در نتیجه تابع خطا در مجموعه داده‌ی آموزشی سریع همگرا می‌شود و پایدار به نظر می‌رسد. در مورد دقت شکل ۱۴، هر انگشت به نرخ دقت مشابهی در پایان آموزش دست یافتند و به حداکثر دقت ۱۰۰ درصد رسیدند. این نتایج با هزینه محاسباتی پایین و زمان کم حاصل می‌شود.

مقدار بالاتری از داده‌های آموزشی احتمالاً نتایج واضح‌تر و معنی داری را ارائه می‌دهد. با این حال، سیستم موفق به گرفتن اشیای مختلف و شیء مورد استفاده در آموزش شد. اشیای دیگر به منظور تست شبکه نیز مورد آزمایش قرار گرفتند تا تعمیم سیستم ارزیابی شود. اشیای دیگر شامل یک هایلاتر، رول چسب کاغذی، تیوپ کرم و تخم مرغ پلاستیکی می‌باشند. این مجموعه در شکل ۱۵ آورده شده است.

تکنیک یادگیری انتقالی اتخاذ شده و از یک VGG-16 از پیش آموزش دیده به عنوان مدل انتخابی برای پیش بینی استفاده از هر انگشت به منظور گرفتن اشیاء استفاده می‌شود. برای آموزش شبکه، وزن لایه‌های اولیه ثابت نگه داشته می‌شود و فقط لایه‌های کاملاً متصل با داده‌های جمع آوری شده دوباره آموزش و تنظیم می‌شوند. شبکه با ۳۰ دوره‌ی آموزشی و نرخ یادگیری ۰/۰۰۰۵ آموزش داده می‌شود. پس از آموزش هر انگشت توسط ۵ شبکه، به منظور بهبود عملکرد و روابط بین انگشتان، شبکه‌ها در ۱ شبکه ترکیب می‌شوند. از چارچوب تنسورفلو به عنوان بستری برای یادگیری عمیق استفاده می‌شود که مجموعه داده‌ها، مدل‌ها و ابزارهای کاربردی را برای وظایف بینایی کامپیوتر فراهم می‌کند. کد مورد استفاده به زبان پایتون نوشته شده است که شبکه‌های از پیش آموزش دیده، دوربین و آردوینو را در یک برنامه کاربردی ترکیب می‌کند. سپس فرمان «دست باز» به آردوینو ارسال می‌شود. پس از باز شدن دست با استفاده از حرکت سروو موتورها، دوربین شروع به پخش فریم به صفحه‌ی کامپیوتر می‌کند. این کار اجازه می‌دهد تا یک شیء قبل از فراخوانی فرمان "دست بسته" به موقعیت مورد نظر منتقل شود. در نهایت خروجی شبکه به یک پیش بینی "استفاده شده" و "استفاده نشده" برای هر انگشت تبدیل می‌شود. به عنوان مثال خروجی ۱۱۰۰۰ به معنای استفاده شدن از انگشت شست و اشاره و استفاده نشدن از سه انگشت دیگر می‌باشد. سپس دست با استفاده از انگشتان فعال شده و فرمان بسته شدن، گرفتن را انجام می‌دهد. این سیستم پس از آموزش و یادگیری، با انتخاب‌های دقیق خود به حداقل ورودی انسانی نیاز پیدا می‌کند و موجب کاهش چشمگیری در نیروی انسانی می‌شود. شکل ۱۲ مراحل گرفتن اشیاء با استفاده از دست ربات را نشان می‌دهد.

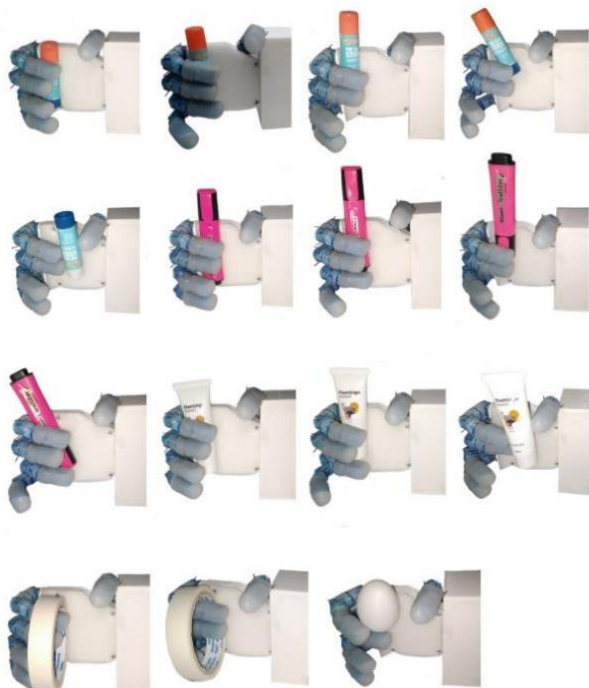


شکل ۱۲) مراحل گرفتن اشیاء

۵- بحث و نتایج

این بخش عملکرد الگوریتم یادگیری و گرفتن در شرایط مشابه را ارائه می‌دهد. در این مقاله از شبکه عصبی کانولوشن برای پیش‌بینی عمل گرفتن بر روی تصاویر یک چسب ماتیکی استفاده شد. شکل‌های ۱۳ و ۱۴ به ترتیب نمودارهای خطا و دقت شبکه برای هر انگشت را نشان می‌دهند. برای خطای شبکه از تابع خطای هوبر استفاده شده است [28].

همانطور که در جدول ۳ مشاهده می‌شود این مطالعه به نتایج قابل قبولی دست یافته است. در شکل ۱۶ چند نمونه از تلاش‌های موفق برای گرفتن چسب ماتیک و سایر اشیاء آورده شده است. جمع آوری نمونه‌های بیشتر برای آموزش شبکه نیز نتایج را بهبود می‌بخشد.



شکل ۱۶) نمونه‌هایی از تلاش‌های موفق برای گرفتن چسب ماتیک، هایلاپتر، تیوپ کرم، رول چسب کاغذی و تخم مرغ پلاستیکی

۶- نتیجه گیری

در این مقاله یک دست ربات ۵ انگشتی ارائه شد که از دوربین و سروو موتور برای جمع‌آوری داده استفاده می‌کند و با استفاده از یک شبکه عصبی کانولوشن آموزش می‌بیند. استفاده از شبکه پیشنهادی از پیش آموزش دیده منجر به کاهش در محاسبات و زمان شد. هدف این مقاله پیاده سازی یک الگوریتم یادگیری عمیق برای پیش بینی استفاده از هر انگشت در گرفتن اشیاء می‌باشد. پس از آموزش، عمل گرفتن دست رباتیک بر روی چند شیء دیگر با اشکال و اندازه‌های مختلف تست شد. نتایج نشان داد که بالاترین میزان موفقیت برای گرفتن یک شیء استوانه‌ای با دقت ۸۵ درصد می‌باشد. هم چنین مشاهده شد شبکه در گرفتن اشیاء ناشناخته هم عملکرد قابل قبولی دارد. دست ربات پیشنهادی می‌تواند برای استفاده در پروتوزهای مختلف سازگار شود. از کارهای آینده می‌توان به بررسی رویکرد عمل گرفتن با استفاده از یادگیری تقویتی عمیق نیز اشاره نمود. اگرچه روش‌های یادگیری عمیق دارای قابلیت تعمیم بالایی هستند اما استفاده از الگوریتم‌های یادگیری تقویتی ممکن است نتایج بهتری را در عمل گرفتن ارائه دهند زیرا عامل به هیچ دانش قبلی در مورد ربات نیاز



شکل ۱۵) مجموعه اشیاء برای آموزش و تست

گرفتن موفق به این صورت تعریف می‌شود که دست ربات بتواند شیء را حداقل به مدت ۲ ثانیه به‌طور ثابت نگه دارد و اگر نتواند شیء را به‌طور ثابت نگه دارد، گرفتن ناموفق تلقی می‌شود. جدول ۲ نشان دهنده درصد موفقیت و شکست برای هر شیء با بیش از ۱۰۰ تلاش برای گرفتن و ۳۰ دوره‌ی آموزشی می‌باشد که به طور خلاصه آورده شده است.

جدول ۲) درصد موفقیت و شکست برای هر شیء

شیء	درصد موفقیت	درصد شکست
چسب ماتیک	۸۵	۱۵
هایلاپتر	۸۳	۱۷
تیوپ کرم	۸۰	۲۰
رول چسب کاغذی	۶۲	۳۸
تخم مرغ پلاستیکی	۵۱	۴۹

جدول ۲ نشان می‌دهد که گرفتن چسب ماتیک با میزان موفقیت ۸۵ درصد دارای بالاترین دقت و گرفتن تخم مرغ پلاستیکی با میزان موفقیت ۵۱ درصد دارای کمترین دقت نسبت به سایر اشیاء می‌باشد. اشیای دیگر مانند هایلاپتر و تیوپ کرم از مجموعه اشیای دیده نشده قبلی به علت شباهت شکل آن‌ها به چسب ماتیک تقریباً به همان اندازه خوب عمل می‌کنند. عملکرد تخم مرغ پلاستیکی در مقایسه با سایر اشیاء به دلیل ناتوانی دست ربات در عمل گرفتن شعاعی ضعیف می‌باشد، در نتیجه گرفتن اشیای کروی به دلیل انحنای سطحی چالش برانگیز است. به منظور تأیید بیشتر اثربخشی مطالعه پیشنهادی، این نتایج با مطالعات گزارش شده قبلی مقایسه شد. در جدول ۳ مقادیر دقت گرفتن تحقیق حاضر و مقالات اخیر آورده شده است.

جدول ۳) مقایسه‌ی درصد گرفتن موفق

دست	درصد گرفتن موفق	کارهای مشابه
گریپر چند انگشتی	۳۰-۶۰	[۳۰]
دست ربات Boris	۷۷/۸	[۳۱]
دست ربات پیشنهادی	۸۵	پژوهش حاضر

networks. *Communications of the ACM*. 2017;60(6):84-90.

18- Bishop CM. *Neural networks and their applications*. Review of scientific instruments. 1994;65(6):1803-32.

19- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014:1409.1556.

20- [Online]. <https://www.geeksforgeeks.org/vgg-16-cnn-model/>

21- Gómez G, Hernandez A, Eggenberger Hotz P, Pfeifer R. An adaptive learning mechanism for teaching a robotic hand to grasp. In *International symposium on adaptive motion of animals and machines 2005*.

22- Levine S, Pastor P, Krizhevsky A, Ibarz J, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International journal of robotics research*. 2018;37(4-5):421-36.

23- Devaraja RR, Maskeliūnas R, Damaševičius R. Design and evaluation of anthropomorphic robotic hand for object grasping and shape recognition. *Computers*. 2020;10(1):1.

24- McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943; 5:115-33.

25- Stergiou C, Siagnos D. *Surveys and Presentations in Information Systems Engineering*. SURPRISE 96 Journal., 2006: p. 1-25.

26- Yoo HJ. Deep convolution neural networks in computer vision: a review. *IEIE Transactions on Smart Processing and Computing*. 2015;4(1):35-43.

27- Liu M, Shi J, Li Z, Li C, Zhu J, Liu S. Towards better analysis of deep convolutional neural networks. *IEEE transactions on visualization and computer graphics*. 2016;23(1):91-100.

28- *Mathworks: Introducing Deep Learning with MATLAB*. 2017.

29- [Online]. https://en.wikipedia.org/wiki/Huber_loss.

30- Kootstra G, Popović M, Jørgensen JA, Kuklinski K, Miatliuk K, Kragic D, Krüger N. Enabling grasping of unknown objects through a synergistic use of edge and surface information. *The international journal of robotics research*. 2012;31(10):1190-213.

31- Kopicki M, Detry R, Adjigble M, Stolkin R, Leonardis A, Wyatt JL. One-shot learning and generation of dexterous grasps for novel objects. *The International Journal of Robotics Research*. 2016 ;35(8):959-76.

ندارد و آموزش فقط با پاداش مثبت یا منفی صورت می‌گیرد. این روش هم چنین برای آموزش ربات‌هایی با درجات آزادی بالا و در محیط‌های بدون ساختار بسیار مطلوب است.

منابع

- 1- Schwarz RJ, Taylor C. The anatomy and mechanics of the human hand. *Artificial limbs*. 1955;2(2):22-35.
- 2- Wheatland N, Wang Y, Song H, Neff M, Zordan V, Jörg S. State of the art in hand and finger modeling and animation. In *Computer Graphics Forum 2015 (Vol. 34, No. 2, pp. 735-760)*.
- 3- Iberall AR. *A neural model of human prehension*. University of Massachusetts; 1987.
- 4- Ansuini C, Santello M, Massaccesi S, Castiello U. Effects of end-goal on hand shaping. *Journal of neurophysiology*. 2006;95(4):2456-65.
- 5- Cutkosky MR. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on robotics and automation*. 1989;5(3):269-79.
- 6- Bicchi A, Marigo A. Dexterous grippers: Putting nonholonomy to work for fine manipulation. *The International Journal of Robotics Research*. 2002;21(5-6):427-42.
- 7- Armada MA, de González Santos P, Alba D, Armada M, Ponticelli R. An introductory revision to humanoid robot hands. In *Climbing and Walking Robots: Proceedings of the 7th International Conference CLAWAR 2004 2005 (pp. 701-712)*. Springer Berlin Heidelberg.
- 8- [Online]. Available: <https://www.openbionics.com/shop/ada>.
- 9- [Online]. Available: <https://inmoov.fr/>
- 10- Mouri T, Endo T, Kawasaki H. Review of gifu hand and its application. *Mechanics based design of structures and machines*. 2011;39(2):210-28.
- 11- Butterfaß J, Grebenstein M, Liu H, Hirzinger G. DLR-Hand II: Next generation of a dextrous robot hand. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164) 2001 (Vol. 1, pp. 109-114)*. IEEE.
- 12- Fischinger D, Weiss A, Vincze M. Learning grasps with topographic features. *The International Journal of Robotics Research*. 2015;34(9):1167-94.
- 13- Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*. 2015;34(4-5):705-24.
- 14- Nair V, Hinton GE. 3D object recognition with deep belief nets. *Advances in neural information processing systems*. 2009;22.
- 15- LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*. 1995;3361(10):1995.
- 16- Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA transactions on Signal and Information Processing*. 2014;3: e2.
- 17- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural