



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی

هدایت پهباد با علائم دست مبتنی بر بینایی ماشین

نگارش
سارا تاجرنیا

استاد راهنما
دکتر مهدی جوانمردی

اردیبهشت ۱۴۰۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی

هدایت پهباد با علائم دست مبتنی بر بینایی ماشین

نگارش
سارا تاجرنیا

استاد راهنما
دکتر مهدی جوانمردی

اردیبهشت ۱۴۰۳

صفحه فرم ارزیابی و تصویب پایان نامه - فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تأیید و تصویب پایان نامه موسوم به فرم کمیته دفاع - موجود در پرونده آموزشی - را قرار دهید.

نکات مهم:

- نگارش پایان نامه/رساله باید به **زبان فارسی** و بر اساس آخرین نسخه دستورالعمل و راهنمای تدوین پایان نامه های دانشگاه صنعتی امیرکبیر باشد.(دستورالعمل و راهنمای حاضر)
- رنگ جلد پایان نامه/رساله چاپی کارشناسی، کارشناسی ارشد و دکترا باید به ترتیب مشکی، طوسی و سفید رنگ باشد.
- چاپ و صحافی پایان نامه/رساله بصورت **پشت و رو(دورو)** بلامانع است و انجام آن توصیه می شود.



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

به نام خدا

تعهدنامه اصالت اثر

تاریخ: اردیبهشت ۱۴۰۳

اینجانب **سارا تاجرنیا** متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیرکبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است.

در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه متعلق به دانشگاه صنعتی امیرکبیر می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

سارا تاجرنیا

امضا

این پایان نامه را تقدیم می‌کنم به مهرباترین همراهان زندگیم، پدر، مادر، برادران
عزیزم که حضورشان همیشه گرمابخش روح من بوده است.

سپاسگزاری

زندگی دفتري از خاطره هاست، يك نفر در دل شب، يك نفر در دل خاك، يك نفر همدم خوشبختي هاست، يك نفر همسفر سختي هاست، چشم تا باز كنيم، عمرمان مي گذرد ما همه رهگذريم، آنچه باقيست فقط خوبي هاست.

تشكر مي كنم از تمامي عزيزاني كه در تمامي مراحل زندگي همراه من بوده اند.
و همچنين از استاد گرامي جناب آقاي دكتر مهدي جوانمردى كه در انتخاب و پيشبرد اين پروژه به عنوان استاد پروژه، كمك هاي فراواني به اين جانب داشتند، كمال تشكر را دارم.

سارا تاجرنا
اردیبهشت ۱۴۰۳

چکیده

پهپادهای تجاری که به عنوان هواپیماهای بدون سرنشین^۱ نیز شناخته می‌شوند، به سرعت در حال رایج شدن هستند و در بسیاری از کاربردهای مختلف مانند نظارت برای رویدادهای ورزشی، حمل و نقل تجهیزات و کالاهای اضطراری، فیلمبرداری، عکاسی هوایی و بسیاری از فعالیتهای دیگر مورد استفاده قرار می‌گیرند.

هدف این پروژه توسعه سیستمی است که از حرکات دست به عنوان روشی برای کنترل پرواز پهپاد استفاده شود. بدین صورت که با استفاده از روش‌های بینایی ماشین^۲، روشی بصری برای ارتباط بدون عامل بین پهپاد و اپراتور آن ایجاد می‌کنیم. روش‌های مبتنی بر بینایی ماشین بر توانایی دوربین هواپیماهای بدون سرنشین متکی هستند. بدین صورت که تصاویر اطراف را گرفته و با استفاده از ترجمه تصاویر و تشخیص الگوی دست، اطلاعات معناداری را استخراج می‌کنند. ساختار این پروژه از دو ماژول اصلی تشکیل شده است: تشخیص حرکت دست^۳ و دستور به هواپیمای بدون سرنشین. برای ماژول اول از یک روش یادگیری عمیق^۴ استفاده شده است. الگوریتم‌ها و تکنیک‌های پردازش تصویر به عنوان روشی پویا برای شناسایی ژست‌ها و حرکات دست معرفی شده‌اند. ماژول دوم وظیفه ارتباط با پهپاد را بر عهده دارد. بدین صورت که پیام‌های بین سیستم پیشنهادی و پهپاد متصل به سیستم را ارسال و دریافت می‌کند و طبق آن پیام‌ها عملیات مورد نظر را اجرا می‌کند.

واژه‌های کلیدی:

پهپاد، هواپیمای بدون سرنشین، ژست دست، بینایی ماشین، شبکه‌های عصبی پیچشی^۵، حافظه طولانی

¹Unmanned aerial vehicles

²Computer vision

³Hand detection

⁴Deep learning

⁵Convolutional neural network

⁶Long short-term memory

⁷Machine learning

⁸Human-drone interface

کوتاه مدت^۶، یادگیری ماشین^۷، رابط انسان و پهباد^۸

فهرست مطالب

صفحه

عنوان

۱	۱ مقدمه
۲	۱-۱ مقدمه
۳	۲-۱ چالش‌های استفاده از پهپاد
۳	۳-۱ اهمیت استفاده از بینایی ماشین در پهپاد
۵	۴-۱ تعریف مسئله
۵	۱-۴-۱ چالش‌های اجرای پروژه
۵	۵-۱ مراحل انجام پروژه
۶	۶-۱ جمع بندی
۷	۲ کارهای مشابه
۸	۱-۲ مقدمه
۹	۲-۲ مقالات مربوط به ویژگی‌های تصویر
۹	۱-۲-۲ مقاله Hand Gesture Controlled Drones: An Open Source Library
۱۰	۲-۲-۲ مقاله Hand Gesture Controlled Drones: An Open Source Library
	۳-۲-۲ مقاله Deep Learning Based Hand Gesture Recognition and UAV
۱۱	Flight Controls
۱۲	۳-۲ مقالات مربوط به ورودی تصویر دست
۱۲	۱-۳-۲ مقاله UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition
۱۳	۴-۲ مقالات مربوط به نقاط کلیدی دست
	۱-۴-۲ مقاله Ap- و MediaPipe Hands: On-device Real-time Hand Tracking
	plying Hand Gesture Recognition for User Guide Application Using
۱۳	MediaPipe
۱۴	۵-۲ مقالات مربوط به اجرای مدل‌های بینایی کامپیوتر روی پهپاد
۱۴	۶-۲ جمع بندی
۱۵	۳ روش انجام پروژه

۱۶	۱-۳ مقدمه
۱۶	۲-۳ دیتاست
۱۶	۳-۳ مدیاپایپ
۱۷	۱-۳-۳ مدل تشخیص کف دست
۱۸	۲-۳-۳ مدل تشخیص نقاط عطف دست
۱۹	۴-۳ اهمیت ژست دست
۲۰	۵-۳ کنترل پهپاد
۲۱	۶-۳ ابزار ها و نرم افزار های مورد استفاده
۲۱	۱-۶-۳ TensorFlow
۲۱	۲-۶-۳ MediaPipe

فهرست تصاویر

شکل	صفحه
۱-۲	چارچوب کنترل پهنپاد مبتنی بر ژست ۱۰
۲-۲	ویژگی های Haar برای استفاده از آستانه رنگ پوست برای تشخیص دست ۱۰
۳-۲	معماری VGG-16 ۱۱
۱-۳	نمونه‌ای از ژست‌های انتخاب‌شده در مجموعه داده‌ها ۱۷
۲-۳	معماری مدل آشکارساز کف دست ۱۸
۳-۳	معماری مدل نقطه عطف دست. این مدل دارای سه خروجی است که یک استخراج کننده ویژگی را به اشتراک می گذارند. هر سر توسط مجموعه داده های مربوطه که با همان رنگ مشخص شده اند آموزش داده می شود. ۱۹
۴-۳	موقعیت ۲۱ نقطه کلیدی در ناحیه دست ۱۹

صفحه	فهرست جداول	جدول
------	-------------	------

فصل اول

مقدمه

۱-۱ مقدمه

پهپادها یا به عبارتی هواپیماهای بدون سرنشین امروزه در صنایع مختلف به عنوان یک فناوری بسیار گسترده و کارآمد مورد استفاده قرار می‌گیرند. هواپیماهای بدون سرنشین اساساً به عنوان ربات‌های پرنده‌ای دیده می‌شوند که عملکردهای متعددی مانند جمع‌آوری داده‌ها و سنجش از محیط اطراف را بر عهده دارند [۱]. از جمله این صنایع می‌توان به کشاورزی، ساخت و ساز، خدمات حمل و نقل و نقشه‌برداری اشاره کرد. یکی از دلایل اصلی افزایش کاربرد این هواپیماهای بدون سرنشین، کارایی بالای آنها است. این فناوری نه تنها به دلیل سرعت بالا در پوشش‌دهی مساحت‌های گسترده، بلکه به دلیل قابلیت برنامه‌ریزی و استفاده در صنایع مختلف مورد توجه قرار می‌گیرد. همچنین، صرفه‌جویی در هزینه‌های مالی و جانی و افزایش امنیت نیز از جمله عوامل مهمی است که اهمیت پهپادها را بیشتر می‌کند [۱].

در حال حاضر، ربات‌های پرنده در مشاغل مختلف مانند سیستم‌های تحویل بسته استفاده می‌شوند [۱]. به عنوان مثال، شرکت‌هایی مانند آمازون و UPS از پهپادهای چند روتور برای تحویل بسته‌های خود استفاده می‌کنند [۱]. در پی این موضوع، بسیاری از شرکت‌های تولیدکننده پهپاد تشویق شدند تا انواع مختلفی از ویژگی‌های نرم‌افزاری و سخت‌افزاری مانند حسگرها را به پهپادها اضافه کنند، که ابتدایی‌ترین آنها دوربین است. دوربین بصری یک حسگر ضروری برای پهپادهای فعلی است. هزینه کم، قدرت کم، اندازه کوچک ضبط تصویر، و دستگاه‌های پخش جریان، آنها را به پهپادهای کاربردی و متعدد در بازار تبدیل می‌کند [۱]. در ادامه زمینه مطالعاتی جدیدی به نام رابط هواپیماهای بدون سرنشین و انسان^۱ گشوده شد تا تعامل بین پهپاد و انسان را پیشرفت دهد، که این تعامل مجموعه دستگاه‌های سنتی مانند کنترلر رادیویی^۲ تا کنترل پهپادها با استفاده از وضعیت بدن و دست انسان را شامل می‌شود [۱].

یکی از رویکردهای مورد استفاده برای افزایش کاربرد و دسترسی به پهپادها، استفاده از بینایی ماشین است. این ویژگی معمولاً از طریق پردازش تصویر و با استفاده از شبکه‌های عصبی به کار می‌رود. پهپادهایی که با مدل‌های بینایی ماشین آموزش می‌بینند، توانایی تحلیل تصاویر و ویدئوهای که از محیط اطراف دریافت می‌کنند را دارا هستند. این قابلیت به پهپاد این امکان را می‌دهد که بدون نیاز به تداخل انسانی، وظایفی همچون امنیت، ارسال کالا، پست و این چنین موارد را انجام دهد [۱]. می‌توان گفت هدف اصلی استفاده از بینایی ماشین در پهپادها برای به حداقل رساندن دخالت انسان به صورت مستقیم است. این

¹Human drone interface

²Radio Controller

امر پهپاد را قادر می‌سازد تا تشخیص اشیاء، تشخیص چهره، تحلیل تصاویر، شناسایی الگوهای مختلف و مواردی از این دست را به صورت خودکار انجام دهند [۴].

۲-۱ چالش‌های استفاده از پهپاد

استفاده از پهپادها، با چالش‌های متعددی همراه است. یکی از این چالش‌ها، محدودیت زمان پرواز است که پس از مدتی نیاز به شارژ مجدد دارند. همچنین، محدودیت‌های محیطی نیز می‌تواند به چالش‌هایی بدل شوند؛ زیرا پهپادها به شرایط محیطی مانند آب و هوا، یا وزن و ارتفاع حساس هستند و این موارد می‌تواند در طراحی آنها تأثیر به‌سزایی داشته باشد. در ادامه باید به میزان اهمیت امنیت اطلاعات هم اشاره کرد، زیرا پهپادها به دلیل استفاده از سیستم‌های موقعیت‌یاب و ارتباطات بی‌سیم ممکن است در برابر حملات سایبری آسیب‌پذیر باشند و اطلاعات مهمی که توسط آنها مخابره می‌شود، در معرض خطر قرار گیرد.

همچنین می‌توان به برخی چالش‌هایی که ما هم در این پروژه به صورتی با آنها سر و کار داریم و در تلاشیم آنها را از بین ببریم یا کمتر کنیم اشاره کرد. مانند انتقال اطلاعات، زیرا برای ارتباط با پهپادها از شبکه‌های بی‌سیم استفاده می‌شود و در شرایطی مانند اشباع شبکه یا فاصله بین پهپاد و کنترل‌کننده، ممکن است این ارتباط دچار مشکل شود. علاوه بر این، محدودیت محاسباتی پهپاد نیز با توجه به اهدافی که برای آن در نظر گرفته شده می‌تواند چالش برانگیز باشد؛ زیرا پهپادها به دلیل محدودیت‌های سخت‌افزاری و نرم‌افزاری، دارای پردازشگرها و حافظه‌های محدودی هستند [۴]. قابل ذکر است که با ادامه پیشرفت فناوری پهپاد، می‌توان انتظار داشت که ویژگی‌های جدید و نوآورانه‌ای برای از بین بردن این محدودیت‌ها و چالش‌ها به پهپادهای آینده اضافه شود.

۳-۱ اهمیت استفاده از بینایی ماشین در پهپاد

طبق اعلام پیش‌بینی اداره هوانوردی فدرال، بازار هواپیماهای بدون سرنشین تا سال ۲۰۲۵ به ۱۷ میلیارد خواهد رسید و ۷ میلیون هواپیمای بدون سرنشین به آسمان پرواز خواهند کرد. پهپادهای کنترل از راه دور به تدریج به دستگاه‌های نیمه خودکار یا کاملاً خودکار تبدیل می‌شوند که از پیاده‌سازی مبتنی بر هوش مصنوعی بهره می‌برند. در این پروژه هدف ما هدایت پهپاد با استفاده از علائم دست مبتنی بر بینایی ماشین است که یک حوزه پژوهشی مهم در ترکیب هوش مصنوعی و رباتیک است. استفاده از

حرکات دست در کنترل هواپیماهای بدون سرنشین در حال تبدیل شدن به یک روش محبوب برای تعامل است. این پایان نامه یک سیستم کامل برای کنترل هواپیماهای بدون سرنشین با استفاده از حرکات دست پیشنهاد می‌کند. سیستم پیشنهادی باید در زمان واقعی^۳ کار کند و دقت^۴ خوبی داشته باشد تا بتواند به بهترین نحو ممکن پهپاد را کنترل کند [۵].

در این روش، از سیستم بینایی ماشین به منظور تشخیص و تحلیل حرکات دست از روی تصاویر ویدئویی پهپاد استفاده می‌شود. با استفاده از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی، سیستم قادر است علائم و حرکات دست را تشخیص داده و به تفسیر آنها بپردازد. سپس، براساس تحلیل این حرکات، دستورات مربوطه برای حرکت و کنترل پهپاد را صادر کند. بدین صورت این روش نه تنها از دقت بالا برای تشخیص و تفسیر حرکات دست برخوردار است، بلکه قابلیت ارائه یک رابط کاربری بین انسان و پهپاد را نیز فراهم می‌کند. به طوری که با استفاده از حرکات دست کاربر قادر است به راحتی و بدون نیاز به دستگاه‌های کنترل خارجی، پهپاد را هدایت کند [۵].

استفاده از حرکات دست برای کنترل پهپاد مزایای زیادی دارد. ابتدا باید گفت که حرکات دست یک شکل طبیعی ارتباطی هستند و استفاده از آنها برای کنترل پهپاد یک روش شهودی و طبیعی برای تعامل با فناوری است. این امر باعث می‌شود که کاربران بتوانند به راحتی و با کمترین تلاش پهپاد را کنترل کنند. استفاده از حرکات دست به کاربر اجازه می‌دهد پهپاد را با سرعت و دقت بیشتری کنترل کند و محدودیت‌های مرتبط با دستگاه‌های کنترل سنتی را کاهش دهد. همچنین، این روش، حرکت و دنبال کردن پهپاد را آسان‌تر می‌کند و امکان جابجایی پهپاد در فضا را بهبود می‌بخشد.

استفاده از علائم دست سبب کاهش نیاز به دستگاه‌های کنترل پیچیده می‌شود و به این ترتیب، پهپاد را برای طیف وسیع‌تری از کاربران قابل دسترس می‌کند. این امر به کاربرانی که با دستگاه‌های کنترل سنتی آشنایی ندارند، امکان استفاده آسان از پهپاد را می‌دهد. همچنین، با توجه به چالش‌هایی که از قبل بیان شده است، این روش خطرات مرتبط با اتصالات بی‌سیم بین کنترلر و پهپاد را کاهش می‌دهد و دقت در کنترل پهپاد در محیط‌های پرتلاطم و مختلف را افزایش می‌دهد.

³Real-time

⁴Accuracy

⁵Image Processing

⁶Deep Neural Network

⁷Convolutional Neural Network(CNN)

۴-۱ تعریف مسئله

هدف این پروژه کنترل کردن پهباد با استفاده از پردازش تصویر^۵ در زمان واقعی است. برای پیاده‌سازی آن می‌توان از یک شبکه عصبی عمیق^۶، مانند یک شبکه عصبی کانولوشن^۷، استفاده کرد. دلیل استفاده از این معماری قابلیت استخراج خودکار ویژگی‌ها با توجه به الگوریتم طبقه‌بندی تصاویر^۸ است. عملکرد شبکه عصبی کانولوشنال به این گونه است که ویژگی‌ها را با توجه به لایه‌های پنهان می‌آموزد، همچنین می‌تواند تعداد پارامترها را بدون به خطر انداختن دقت مدل تغییر دهد. با گذشت زمان محققان معماری‌های مختلفی از شبکه عصبی کانولوشن را برای دقت^۹ بهتر، زمان پردازش کمتر و پیچیدگی‌های^{۱۰} گوناگون مطرح کردند.

۱-۴-۱ چالش‌های اجرای پروژه

وجود سخت‌افزاری مناسب برای اجرای این پروژه الزامی است. پهباد انتخاب شده در ابتدا باید شامل یک دوربین با رزولوشن نسبتاً بالا (حداقل **** پیکسل باشد) تا ژست دست تا فاصله سه متری از پهباد به وضوح گرفته شود. در ادامه از آنجایی که زمان واقعی در این پروژه از اهمیت بالایی برخوردار است پهباد باید پردازنده نسبتاً قوی داشته باشد تا بتواند به صورت مستقل و بدون نیاز به هیچ‌گونه سخت افزار خارجی مدل را اجرا کند، بدین صورت که در هر لحظه ورودی عکس گرفته‌شده از دوربین را به مدل بدهد و در کمترین زمان ممکن بتواند خروجی مدل را به دست آورده و دستور مورد نظر را روی پهباد به اجرا درآورد. از دیدگاهی دیگر، از آنجایی که اجرای یک مدل بینایی ماشین یک برنامه سنگین است و اجرای آن برای عموم پهبادهای انرژی زیادی می‌طلبد، لذا باید پهبادی را انتخاب کرد که از شامل باتری بادوام و باکیفیت باشد که هم در هنگام اجرای مدل بتواند انرژی موردنیاز پردازنده را فراهم کند و همچنین عمر کوتاه آن به مرور زمان برای استفاده کننده آزاردهنده نباشد.

۵-۱ مراحل انجام پروژه

۱. پیاده‌سازی کدی برای جمع‌آوری دیتاست

۲. پیاده‌سازی مدل‌های مدنظر

^۸Image Classification

^۹Accuracy

^{۱۰}Complexity

۳. آموزش به مدل‌ها

۴. بهینه‌سازی مدل‌ها

۵. تست مدل‌ها و انتخاب بهترین مدل

۶. اجرای مدل روی پهنپاد

۱-۶ جمع بندی

هدف این پروژه پیاده‌سازی برنامه‌ای کاربردی بر روی پهنپاد است تا بتواند ۹ ژست دست از پیش تعیین شده را شناسایی و با توجه به آنها دستور داده شده از طرف کاربر را به پهنپاد بدهد. در این پروژه موارد زیر از اهمیت بالایی برخوردار هستند:

- دقت بالای مدل زیرا در صورت انجام نادرست دستورات امکان برآورد هزینه مالی وجود دارد.
- زمان واقعی زیرا اجرای دستورات باید در کمترین زمان ممکن رخ دهد تا مورد پسند کاربر باشد.
- مدلی سبک تا توانایی اجرا روی پردازنده پهنپاد را داشته‌باشد.

فصل دوم

کارهای مشابه

۱-۲ مقدمه

در این فصل هدف ما بررسی پروژه های مشابه است تا بتوان از آنها در روند پروژه کمک گرفت. همچنین در این راه می توان با توجه به نتایج و ارزیابی پروژه های دیگر بستری را فراهم کرد تا نتیجه پروژه را با دیگر کارهای مشابه مقایسه کرد.

به صورت کلی پروژه هایی با هدف کنترل پهپاد با ژست دست در ۲ دسته قرار می گیرند.

- کنترل پهپاد با کمک بینایی ماشین که شامل شبکه هایی برای پردازش تصویر است.
- کنترل پهپاد با دستکش های سنسور دار از جمله سنسور IMU که نیازمند سخت افزار خاص برای پیدا کردن موقعیت نقاط دست است. مانند پروژه های Motion Estimation and Hand Gesture Recognition-Based Human-UAV Interaction Approach in Real Time [۱] و Hand gesture recognition with convolutional neural networks for the multimodal UAV control [۲].

- وجود دستگاه کنترل کننده حرکت جهشی Leap Motion Controller که با توجه آن ویژگی های دست با دقت بالا اندازه گیری شده و با کمک شبکه های عصبی ژست دست تشخیص داده میشود. پروژه ی Deep Learning Based Hand Gesture Recognition and UAV Flight Controls [۳] و Gesture control of drone using a motion controller [۴] نمونه ای از این جمله پروژه ها هستند.

از بین این موارد پروژه ما مربوط به اولین گزینه است که تنها سخت افزار مورد نیاز به جز پهپاد دوربین نصب شده روی پهپاد است. که به بررسی نمونه ای این پروژه ها می پردازیم. پروژه های مشابه با کار ما که با کمک بینایی ماشین پهپاد را کنترل می کنند به ۴ دسته کلی تفکیک می شوند.

۱. استخراج ویژگی های تصویر در هر فریم که با توجه به نیازهای مسئله می تواند متفاوت باشد.
۲. تشخیص دست^۱ برای پیدا کردن موقعیت دست در هر فریم تصویر و ورودی پیکسل های RGB آن به مدل و در نهایت کلاس بندی ژست دست.
۳. استخراج نقاط کلیدی^۲ تصویر و ورودی آنها به مدل برای کلاس بندی.

^۱Hand detection

^۲Key point

۲-۲ مقالات مربوط به ویژگی‌های تصویر

۱-۲-۲ مقاله Hand Gesture Controlled Drones: An Open Source Library

این پروژه بر پیاده سازی یک سیستم کنترل برای هواپیماهای بدون سرنشین با استفاده از حرکات دست، مشابه رویکرد مورد بحث در مقاله تمرکز دارد. هدف آن استفاده از شبکه‌های عصبی یادگیری عمیق برای تشخیص لحظه‌ای حرکات دست پویا برای کنترل پرواز پهپاد است. بدین صورت که با توجه به سایه‌ها و رنگ‌های درون تصویر ویژگی‌های Haar مورد نظر را پیدا کرده و با کمک آنها ژست دست را تشخیص می‌دهد.

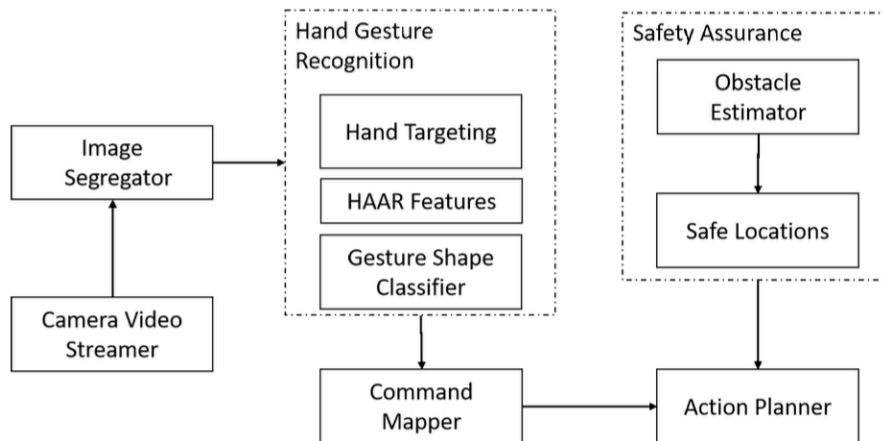
روش‌شناسی

پروژه پیاده‌سازی شده حاوی طراحی و آموزش شبکه‌های عصبی در راستای حرکت پهپاد می‌باشد. این سیستم شامل پیش پردازش داده‌ها، انتخاب ویژگی، ماژول شبکه عصبی یادگیری عمیق برای تشخیص ژست و ماژول کنترل پهپاد برای ترجمه ژست‌های شناسایی شده به دستورات حرکت پهپاد است. برای شناسایی و تشخیص ژست‌های دست، ابتدا از یک شبکه عصبی برای تشخیص موقعیت دست استفاده می‌شود. این شبکه با استفاده از داده‌های ورودی به صورت تصویری، موقعیت دقیق دست را در تصویر تعیین می‌کند. پس از تشخیص موقعیت دست، ویژگی‌های Haar از تصویر استخراج می‌شوند. ویژگی‌های Haar مجموعه‌ای از الگوریتم‌های تشخیص ویژگی است که به صورت یکپارچه و متناوب از تصاویر استفاده می‌کنند تا ویژگی‌های خاصی از تصویر را شناسایی کنند. بدین صورت که مجموعه‌ای از فیلترهای مختلف به عنوان فیلتر Haar به تصویر اعمال می‌شود. این فیلترها از الگوهای ساده مانند خطوط عمودی و افقی یا مربع‌های ساده تشکیل شده‌اند. با اعمال این فیلترها بر روی تصویر، مقادیری به دست می‌آید که نشان‌دهنده تفاوت‌های میانگین شدت نور در مناطق مختلف تصویر است. در نهایت، این ویژگی‌ها به عنوان ورودی یک شبکه SVM^۳ داده شده تا ژست دست تشخیص داده شود.

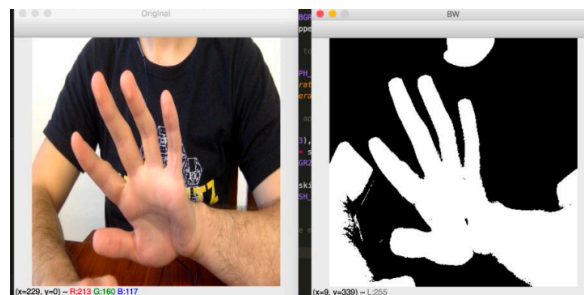
نتیجه بدست آمده

این پروژه دقت بالایی در تشخیص ژست دست و کنترل پرواز پهپاد دست می‌یابد. برای خروجی این پروژه ۵ حالت دست مدنظر قرار گرفته‌اند. دقت متوسط آن برابر ۹۷.۴۷۱ درصد است که عملکرد عالی

^۳Support Vector Machine



شکل ۱-۲ چارچوب کنترل پهباد مبتنی بر ژست



شکل ۲-۲ ویژگی های Haar برای استفاده از آستانه رنگ پوست برای تشخیص دست

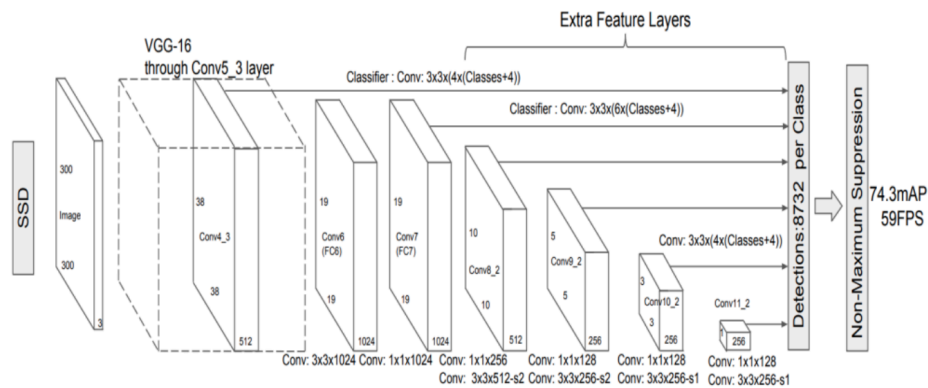
را نشان می‌دهد. البته قابل ذکر است که این دقت در پس‌زمینه‌های به هم ریخته و همچنین در شرایط نوری مختلف بسیار متغیر است زیرا همان‌طور که گفته شد ویژگی Haar به سایه و رنگ‌های درون تصویر بسیار حساس است [۹].

۲-۲-۲ مقاله Hand Gesture Controlled Drones: An Open Source Library

در این پروژه نیز از ویژگی‌های Haar استفاده شده تا بتوان ۹ حالت دست را شناسایی کرد.

روش‌شناسی

در این مقاله نیز مانند پروژه قبل از استخراج ویژگی‌های Haar استفاده شده. با این تفاوت که در مقاله قبل از معماری SVM برای کلاس‌بندی ژست‌های دست استفاده شده است، در حالی که در این مقاله از شبکه VGG-16 که به صورت لایه‌های کانولوشن استفاده شده.



شکل ۲-۳ معماری VGG-16

نتیجه بدست آمده

در پروژه این پروژه ۹ حالت دست مدنظر قرار گرفته شده و دقت آن برابر ۸۳.۳ درصد است. که در بهترین حالت ممکن با پس زمینه‌ی مناسب بدست آمده و باید در نظر گرفت که دقت بالایی برای کنترل پهپاد به حساب نمی‌آید.

۲-۲-۳ مقاله Deep Learning Based Hand Gesture Recognition and UAV

Flight Controls

در این مقاله با استفاده از استخراج ویژگی‌های تصویر و ورودی آنها به یک شبکه بینایی کامپیوتر ژست دست تشخیص داده می‌شود.

روش شناسی

برای پیاده‌سازی این پروژه از هر یک از فریم‌های دریافت شده توسط پهپاد با سنسورهای به خصوصی پارامترهایی از جمله زاویه انحراف، مختصات، قدرت گرفتن دست استخراج می‌شود. آنها به صورت یک ماتریس 15×45 درآمده و در ادامه با ورود این پارامترها به صورت متوالی به شبکه کانولوشن با معماری زیر داده شده و دریافت خروجی می‌تواند هدف حرکت کاربر را نشان دهد.

نتیجه

[؟]

۳-۲ مقالات مربوط به ورودی تصویر دست

۱-۳-۲ مقاله UAV-GESTURE: A Dataset for UAV Control and Gesture

Recognition

این مقاله با هدف کنترل پهپاد یا خلبان خودکار با کمک حرکت دست پیاده‌سازی شده است. برای مثال حرکت چپ به راست دست نشان دهنده حرکت پهپاد به راست است. بدین ترتیب برای اجرای این برنامه شبکه P-CNN طراحی شده است تا بتواند معنای عکس‌ها را تجربه کند.

روش‌شناسی

در این مقاله از شبکه P-CNN استفاده می‌شود. بدین‌صورت که در ابتدا موقعیت دست فرد با کمک جعبه مرزی^۴ پیدا کرده، سپس عکس دست با فیلترهای مناسب وارد شبکه‌ی P-CNN می‌شود تا بتوان ژست دست را پیش‌بینی کرد. در خروجی مدل ۱۳ نوع حرکت وجود دارد تا بتواند معنای آنها را پیش‌بینی کند. این حرکات برخلاف دیگر مقالات کل دست از شانه تا انگشتان و حرکات آنها شامل می‌شود. زیرا هدف اصلی آن دستور دادن به هواپیماهای بزرگ بدون سرنشین در فرودگاه‌ها است. از آنجایی که دستان در تصاویر گرفته‌شده نسبتاً کوچک‌اند، کیفیت تصاویر باید بالا باشد. از طرف دیگر از آنجایی که در این پروژه حرکات دست مدنظر است نه ژست آنها، توالی موقعیت و ژست دست‌هاست که اهمیت دارد. این توالی فریم‌ها سبب می‌شود تا مدل پیچیده و سنگینی داشته باشیم.

نتیجه

دقت مدل در بهترین حالت، با بهترین دیتاست ممکن برابر ۹۱.۹ درصد است که دقت بالایی برای اجرای پروژه است. اما به دلیل پیچیده و سنگین بودن مدل و همچنین وجود سنسورهای علاوه بر دوربین نصب‌شده روی پهپاد، این پروژه یکی از بزرگ‌ترین اهداف پروژه که زمان واقعی بودن آن است دچار چالش می‌شود. [۴]

⁴Bounding box

۴-۲ مقالات مربوط به نقاط کلیدی دست

مقاله ۱-۴-۲ **MediaPipe Hands: On-device Real-time Hand Tracking**

و **Applying Hand Gesture Recognition for User Guide Application Using MediaPipe**

در این مقاله‌ها از کتابخانه MediaPipe استفاده شده‌است تا بتوان ۲۱ نقطه عطف دست را پیدا کرده و در پروژه‌های گوناگون از جمله پیدا کردن ژست دست و افکت‌های AR استفاده کند. ما در پروژه خود از این کتابخانه استفاده می‌کنیم تا بتوانیم مدلی سبک و ساده پیاده کنیم.

روش‌شناسی

در این مقاله به بررسی این کتابخانه پرداخته‌شده و ما از آن استفاده می‌کنیم تا بتوانیم مدلی برای پیش‌بینی ژست دست استفاده کنیم. برای پیاده‌سازی این پروژه از دو شبکه کانولوشن استفاده شده است. شبکه اول که برای پیدا کردن کف دست در تصویر استفاده می‌شود و شبکه دوم که به عنوان ورودی موقعیت عکس دست پیدا شده را دریافت و مختصات ۲۱ نقطه عطف را موقعیت‌یابی می‌کند.

نتیجه

مدل‌های طراحی شده در این مقالات برای تشخیص نقاط عطف دست از دقت ۹۵.۷ درصد برخوردار هستند که دقت بسیار بالایی محاسبه می‌شود. همچنین این مدل به نور و تصویر پس‌زمینه وابسته نیست و دقت متوسط آن در زمینه‌های مختلف اندازه‌گیری شده لذا مدل را کاربردی و مورد پسندتر می‌کند؟ [؟] .

۵-۲ مقالات مربوط به اجرای مدل‌های بینایی کامپیوتر روی

پهپاد

۶-۲ جمع‌بندی

در تمام مقالات بررسی شده، پروژه‌ها به گونه‌ای پیاده‌سازی شده‌اند تا حرکات را به طوری کلاس‌بندی کنند که در خروجی حتما یکی از ژست‌های در نظر گرفته شده انتخاب شود. لذا زمانی که دست در حالتی غیر از آنها قرار دارد، مدل طراحی شده حتما یکی از ژست‌هایی را که به آن شبیه‌تر است را انتخاب می‌کند که این امر می‌تواند برای پیاده‌سازی روی پهپاد واقعی مشکل‌زا باشد و حتی هزینه مالی به ارمغان آورد.

فصل سوم

روش انجام پروژه

۱-۳ مقدمه

برای اجرای این پروژه راه‌های متفاوتی مورد بررسی قرار داده شد تا بتوان بهترین آنها را روی پهباد پیاده‌سازی کرد. نتیجه نهایی پیاده‌سازی ۳ شبکه کانولوشن به صورت پی در پی است. بدین صورت که شبکه اول برای پیدا کردن موقعیت دست به کار گرفته می‌شود. بدین صورت که هر فریم گرفته‌شده از دوربین پهباد پس از تغییر اندازه به یک ماتریس $256 \times 256 \times 3$ به ورودی مدل داده می‌شود و پس از پردازش آن خروجی یک ماتریس $256 \times 256 \times 3$ است که در آن دست را نشان می‌دهد. پس از آن، این ماتریس به مدل دوم داده می‌شود، در این مدل عکس دست به صورت یک ماتریس گرفته‌شده و خروجی آن برابر ۲۱ نقطه سه بعدی عطف دست و شاخص دست (راست یا چپ) است. در ادامه مدل سوم به عنوان ورودی، یک ماتریس 21×2 می‌گیرد که مختصات نقاط طول و عرض هر نقطه عطف دست است چرا که عمق تصویر با توجه به ژست‌های در نظر گرفته شده از اهمیت بالایی برخوردار نیست. و در خروجی پیش‌بینی می‌کند که کدام ژست دست مد نظر کاربر است. با توجه به این پروژه ۹ ژست گوناگون مدنظر قرار گرفته‌شده (کاربر می‌تواند ژست‌های جدیدی اضافه کند)، خروجی شبکه کانولوشن شامل ۱۰ کلاس کلاس است که ۹ کلاس برای ژست‌ها و کلاسی برای زمانی که هیچ کدام از ژست‌ها انتخاب نشده در نظر گرفته‌شده.

۲-۳ دیتاست

۳-۳ مدیاپایپ

برای پیاده‌سازی شبکه‌های تشخیص کف دست و پیدا کردن نقاط عطف دست از مدل‌های از قبل آموزش دیده^۱ کتابخانه مدیاپایپ کمک گرفته‌شده است. مدیاپایپ از یک خط لوله یادگیری ماشین متشکل از چندین مدل که با هم کار می‌کنند استفاده می‌کند: یک مدل تشخیص کف دست^۲ که تصویر را از ورودی می‌گیرد و عکس محدوده دست را به عنوان خروجی دریافت می‌کند و یک مدل تشخیص نقاط عطف دست^۳ که عکس دست را به عنوان ورودی گرفته و مختصات ۲۱ نقطه کلیدی بندهای انگشتان دست را در ناحیه دست تشخیص می‌دهد.

¹Pretrained

²Palm detection model

³Hand landmark model



شکل ۳-۱ نمونه‌ای از ژست‌های انتخاب‌شده در مجموعه داده‌ها

۳-۳-۱ مدل تشخیص کف دست

مدل تشخیص کف دست مدیاپایپ دارای دقت متوسط ۷.۹۵ درصد است که این دقت بالا با استفاده از استراتژی‌های مختلف به دست آمده است. ابتدا، به جای آشکار کردن دست^۴، آشکار کردن کف دست را به مدل آموزش می‌دهند، زیرا پیدا کردن محدود از اجسام سفت و سخت مانند کف دست و مشت بسیار ساده‌تر از تشخیص دست‌ها با انگشتان مفصلی است. علاوه بر این، از آنجایی که کف دست‌ها اشیاء کوچکی هستند، الگوریتم سرکوب غیر حداکثری^۵ که یک تکنیک پس پردازش^۶ است و در تشخیص

^۴hand detector

^۵Non-maximum suppression

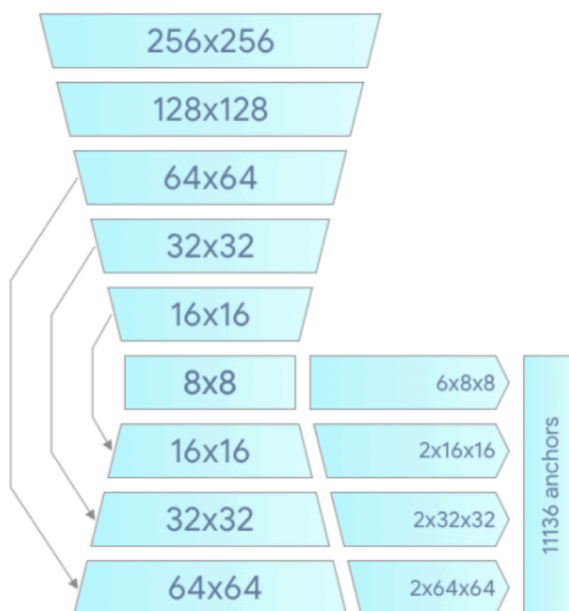
^۶post-process

^۷duplicate detections

^۸false positive

^۹computational complexity

اشیا برای حذف تشخیص های تکراری^۷ و انتخاب مرتبط ترین اشیاء شناسایی شده استفاده می شود. این به کاهش مثبت کاذب^۸ و پیچیدگی محاسباتی^۹ یک الگوریتم تشخیص کمک می کند. تا بهترین محدوده مربعی^{۱۰} با واریانس بالا^{۱۱} را بدست آورد. [۹]



شکل ۲-۳ معماری مدل آشکارساز کف دست

۲-۳-۳ مدل تشخیص نقاط عطف دست

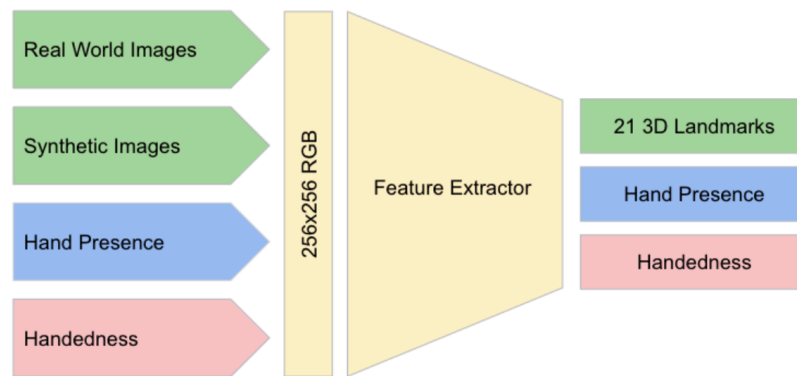
در این مرحله مکان یابی مختصات ۲۱ نقطه کلیدی بندهای انگشتان دست که شامل سه بعد است از طریق رگرسیون^{۱۲} انجام می شود. این مدل بر روی ۳۰ هزار تصویر دنیای واقعی با ۲۱ مختصات سه بعدی برچسب زده شده^{۱۳} آموزش دیده است. برای پوشش بهتر ژست های احتمالی دست و ارائه نظارت بیشتر بر ماهیت هندسه دست، این دیتاست از مدل های دست مصنوعی با کیفیت بالا را نیز روی پس زمینه های مختلف ارائه می کند تا دقت را به بالاترین حد ممکن برساند. این مدل حتی در برابر دست های نیمه نیز عملکرد قوی نشان می دهد. [۹]

¹⁰bounding box

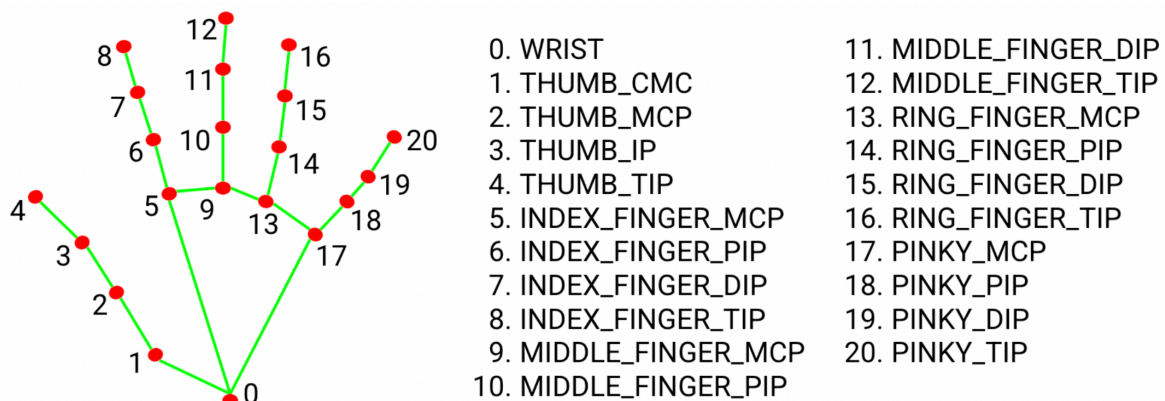
¹¹high scale variance

¹²regression

¹³labeling



شکل ۳-۳ معماری مدل نقطه عطف دست. این مدل دارای سه خروجی است که یک استخراج کننده ویژگی را به اشتراک می گذارند. هر سر توسط مجموعه داده های مربوطه که با همان رنگ مشخص شده اند آموزش داده می شود.



شکل ۳-۴ موقعیت ۲۱ نقطه کلیدی در ناحیه دست

۴-۳ اهمیت ژست دست

وقتی مردم صحبت می کنند، ژست می گیرند. ژست جزء اساسی زبان است که اطلاعات معنادار و منحصر به فردی را انتقال می دهد. ژست ها به گوینده کمک می کنند تا اهداف خود را بهتر منعکس کند. آن ها نقش های بسیاری را در ارتباط، یادگیری و درک هم برای افرادی که آنها را مشاهده می کنند و هم برای کسانی که آنها را ایجاد می کنند، ایفا می کنند. وقتی مردم صحبت می کنند، دستان خود را حرکت می دهند. به حرکات خود به خودی دست که در ریتم گفتار ایجاد می شوند، حرکات هم گفتاری^{۱۴} نامیده می شوند و مردم از همه فرهنگ ها و پیشینه های زبانی شناخته شده ژست می گیرند و برای ارتباط از حرکات هم گفتاری برای رساندن بهتر مفهوم خود کمک می گیرند. در واقع، نوزادان قبل از اینکه اولین کلمات خود را بیان کنند، از انواع ژست ها استفاده می کنند. دستان های ما به ما کمک می کنند صحبت

¹⁴co-speech gestures

کنیم، فکر کنیم، و به خاطر بسپاریم، گاهی دانش منحصر به فردی را که هنوز نمی‌توان به زبان آورد، آشکار می‌کنند. به طوری که می‌توان گفت ژست‌ها اغلب به عنوان زبان گفتاری ثانویه در نظر گرفته می‌شود. [؟] ژست‌ها به‌ویژه زمانی مؤثر هستند که مزیتی نسبت به کلمات داشته باشند. [؟] توانایی درک شکل و حرکت دست‌ها می‌تواند یک جزء حیاتی در بهبود تجربه کاربر^{۱۵} در حوزه‌ها و پلتفرم‌های مختلف فناوری باشد. درک مفهوم ژست دست در زمان واقعی برای افراد به طور طبیعی وجود دارد، یک کار بینایی کامپیوتری کاملاً چالش برانگیز است، زیرا دست‌ها اغلب خود یا یکدیگر را مسدود می‌کنند مانند انسداد انگشت، کف دست و لرزش دست و فاقد الگوهای کنتراست بالا هستند. [؟]

۳-۵ کنترل پهپاد

اکثر پهپادهای تجاری موجود در بازار یا دارای کنترلرهای طراحی شده ویژه هستند، یا دارای فرستنده سیگنال اختصاصی و برنامه‌های نرم‌افزاری هستند که روی دستگاه‌های دستی کاربران مانند تلفن‌های همراه یا تبلت‌ها اجرا می‌شوند. در هر دو مورد، کنترل‌کننده فرمان‌هایی را با اطلاعات دقیق از طریق کانال‌های بی‌سیم مانند وای‌فای یا بلوتوث ارسال می‌کند. اخیراً محصولات تجاری وجود داشته است که حرکات دست را به عنوان یک مکانیسم کنترل قابل اجرا معرفی می‌کنند. برای گرفتن ژست‌ها، دو رویکرد وجود دارد.

- استفاده از دستکش‌های طراحی شده ویژه: کنترل‌کننده بر روی دستکشی که توسط کاربران استفاده می‌شود نصب می‌شود و در زمان واقعی انحراف، گام و چرخش دست را شناسایی می‌کند تا به حرکات مربوطه برای پهپاد را شناسایی و ارسال کند. محصولات عبارتند از Kd Interactive MenKind Motion Control Drone و Aura Drone

- استفاده از بینایی کامپیوتر از طریق دوربین: این دستگاه‌ها از دوربین نصب شده روی پهپاد استفاده می‌کنند تا بتوانند در لحظه تشخیص دهند که دست کاربر کجاست و در چه حالتی قرار دارد تا پهپاد را کنترل کند. محصولات عبارتند از DJI Spark Drone

¹⁵user experience

۳-۶ ابزار ها و نرم افزار های مورد استفاده

۳-۶-۱ TensorFlow

TensorFlow یک کتابخانه نرم افزاری رایگان و منبع باز برای یادگیری ماشین و هوش مصنوعی است. می توان از آن در طیف وسیعی از وظایف استفاده کرد، اما تمرکز ویژه ای بر آموزش و استنتاج شبکه های عصبی عمیق دارد.

۳-۶-۲ MediaPipe

MediaPipe مجموعه ای از کتابخانه ها و ابزارهایی است که از تکنیک های هوش مصنوعی و یادگیری ماشین در برنامه های خود استفاده می کند. این کتابخانه می تواند مطابق با نیازهای افراد خود سفارشی شود و در پلتفرم های مختلف توسعه پیدا کند [۱].

فصل چهارم

نتایج و ارزیابی

۱-۴ مقدمه

۱	۲	۲	۳
۱	۶	۸۷۸۳۷	۷۸۷
۲	۷	۷۸	۵۴۱۵
۳	۵۴۵	۷۷۸	۷۵۰۷
۴	۵۴۵	۱۸۷۴۴	۷۵۶۰
۵	۸۸	۷۸۸	۶۳۴۴

جدول ۱-۴ جدول