



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 جوانمردي كه در انتخاب و پيشبرد اين پروژه به عنوان استاد پروژه، كمك هاي فراواني به اين جانب داشتند، كمال تشكر را دارم.

سارا تاجرنا
ارديشست 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 A real-time hand gesture recognition method
	۳-۲-۲ مقاله Hand Gesture Recognition using Image Processing and Feature Extraction Techniques
۱۲	۳-۲ مقالات مربوط به ورودی تصویر دست به مدل
۱۲	۱-۳-۲ مقاله Hand Gestures For Drone Control Using Deep Learning
۱۳	۲-۳-۲ مقاله UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition
۱۵	۴-۲ مقالات مربوط به نقاط کلیدی دست
۱۵	۱-۴-۲ مقاله Hand Gesture Recognition system for Real-Time Application
	۲-۴-۲ مقاله An improved hand gesture recognition system using keypoints
۱۵	and hand bounding boxes
۱۶	۳-۴-۲ مقاله Visual gesture recognition based on hand key points
۱۶	۴-۴-۲ مقاله MediaPipe Hands: On-device Real-time Hand Tracking
۱۷	۵-۲ مقالات مربوط به اجرای مدل‌های بینایی کامپیوتر روی پهپاد

Modeling relation among implementing AI-based drones and مقاله	۱-۵-۲
..... sustainable construction project success	۱۷
..... جمع‌بندی	۶-۲
..... مقاله	۱-۶-۲
.....	۱۸
۳ روش انجام پروژه	۱۹
..... مقدمه	۱-۳
..... انتخاب ژست‌های دست متناسب با حرکت پهباد	۲-۳
..... دیتاست	۳-۳
..... اهمیت ژست دست	۴-۳
..... کنترل پهباد	۵-۳
..... ابزار ها و نرم افزار های مورد استفاده	۶-۳
..... TensorFlow	۱-۶-۳
..... MediaPipe	۲-۶-۳
..... مدیایایپ	۷-۳
..... مدل تشخیص کف دست	۱-۷-۳
..... مدل تشخیص نقاط عطف دست	۲-۷-۳
..... پیش‌پردازش	۸-۳
..... رأی‌گیری پنجره‌ای	۹-۳
۴ نتایج و ارزیابی	۲۸
..... مقدمه	۱-۴
..... نتایج و ارزیابی	۲-۴
..... جمع‌بندی	۳-۴
۵ نتیجه‌گیری و پیشنهادات	۳۰
..... مقدمه	۱-۵
..... نتیجه‌گیری و پیشنهادات	۲-۵
..... جمع‌بندی	۳-۵

۳۲ کتاب‌نامه
----	-----------------

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	تشخیص کف دست و انگشتان ۱۱
۴-۲	معماری VGG-16 ۱۳
۵-۲	معماری ساختار شبکه های عصبی دو خط لوله ۱۶
۱-۳	نمونه ای از ژست های انتخاب شده در مجموعه داده ها ۲۱
۲-۳	معماری مدل آشکارساز کف دست ۲۵
۳-۳	خط لوله تشخیص دست ۲۵
۴-۳	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. ۲۶
۵-۳	موقعیت ۲۱ نقطه کلیدی در ناحیه دست ۲۶

صفحه	فهرست جداول	جدول
۲۹	۱-۴ 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 [۲].

در حال حاضر، ربات‌های پرنده در مشاغل مختلف مانند سیستم‌های تحویل بسته استفاده می‌شوند [۳]. به عنوان مثال، شرکت‌هایی مانند آمازون و UPS از پهپادهای چند روتور برای تحویل بسته‌های خود استفاده می‌کنند [۴]. در پی این موضوع، بسیاری از شرکت‌های تولید کننده 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 [۵]. 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 [۶].

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 [۷]. می‌توان گفت هدف اصلی استفاده از بینایی ماشین در پهپادها برای به حداقل رساندن دخالت انسان به صورت مستقیم است. این

¹Human drone interface²Radio 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 [۸].

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 [۹]. 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 [۶].

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 [۱۰].

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. انتخاب ژست‌های مناسب و مفهومی برای کنترل پهباد
۲. انتخاب چگونگی ذخیره کردن دیتای عکس‌ها در دیتاست

^۵Image Classification

^۹Accuracy

^{۱۰}Complexity

۳. پیاده‌سازی کد برای جمع‌آوری دیتاست

۴. پیاده‌سازی مدل‌ها متناسب با دیتا

۵. آموزش به مدل‌ها

۶. بهینه‌سازی مدل‌ها

۷. تست مدل‌ها و انتخاب بهترین مدل

۸. پیاده‌سازی رأی‌گیری پنجره‌ای^{۱۱}

۹. اجرای مدل روی پهباد

۶-۱ جمع‌بندی

هدف این پروژه پیاده‌سازی برنامه‌ای کاربردی بر روی پهباد 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 پهباد 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 پهباد را داشته‌باشد.

¹¹Window voting

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 [۱۰] و Hand gesture recognition with convolutional neural networks for the multimodal UAV control [۱۱].

- 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 های Deep Learning Based Hand Gesture Recognition and UAV Flight Controls [۱۲] و Gesture control of drone using a motion controller [۱۳] 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

مقالات به‌کار برده شده در این قسمت بر چگونگی تعیین ژست دست با توجه به تصویر داده‌شده تمرکز دارند. برخی از این مقالات چگونگی ارتباط با پهپاد را نیز پوشش می‌دهند، اما نکته مهم در این مقالات چگونگی استخراج ویژگی‌های تصویر و استفاده از آنها برای تعیین ژست دست است.

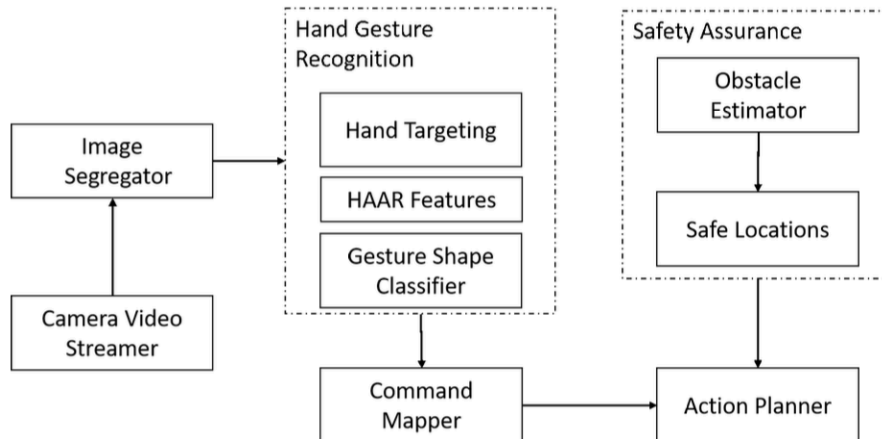
۱-۲-۲ مقاله Hand Gesture Controlled Drones: An Open Source Library

در این پروژه، تمرکز بر پیاده‌سازی یک سیستم کنترل برای هواپیماهای بدون سرنشین با استفاده از حرکات دست است که مشابه رویکرد مورد بحث در مقاله می‌باشد. هدف اصلی این پروژه استفاده از شبکه‌های عصبی یادگیری عمیق برای تشخیص لحظه‌های حرکات دست پویا برای کنترل پرواز پهپاد است. این تشخیص بر اساس ویژگی‌های Haar که با توجه به سایه‌ها و رنگ‌های درون تصویر تعیین می‌شوند، انجام می‌شود.

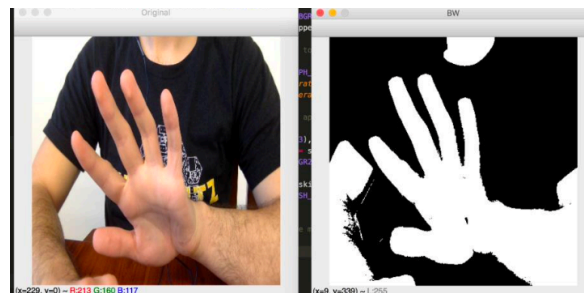
روش‌شناسی

در این پروژه، ابتدا از یک شبکه عصبی برای تشخیص موقعیت دست استفاده می‌شود که به عنوان یک ماژول پیش‌پردازش عمل می‌کند و با دقت بالایی موقعیت دست را تشخیص می‌دهد. پس از تشخیص موقعیت دست، ویژگی‌های Haar از تصویر استخراج می‌شوند که این ویژگی‌ها مجموعه‌ای از الگوریتم‌های تشخیص ویژگی هستند که از تصاویر استفاده می‌کنند تا ویژگی‌های خاصی از تصویر را شناسایی کنند. ویژگی‌های Haar بر اساس تغییرات گرادیان در تصویر تعیین می‌شوند و به عنوان الگوهای محلی برای تشخیص حرکات دست استفاده می‌شوند. در ادامه، از شبکه‌های عصبی یادگیری عمیق برای تشخیص حرکات دست پویا برای کنترل پرواز پهپاد استفاده می‌شود. این شبکه‌ها عملکرد پیچیده‌ای دارند و با استفاده از داده‌های ورودی، مثل ویژگی‌های Haar، می‌آموزند تا حرکات دست را تشخیص دهند و بر اساس آنها دستورات حرکت پهپاد را تعیین کنند. این سیستم شامل مراحل پیش‌پردازش داده‌ها، انتخاب ویژگی، ماژول شبکه عصبی برای تشخیص ژست و ماژول کنترل پهپاد برای ترجمه ژست‌های شناسایی شده به دستورات حرکت پهپاد می‌باشد. علاوه بر این، از مدل SVM (ماشین بردار پشتیبان) برای کلاس‌بندی و تشخیص حرکات دست استفاده شده است. SVM به عنوان یک ماشین یادگیری ماشینی است که برای مسائل دسته‌بندی و رگرسیون استفاده می‌شود و در این پروژه برای تشخیص حرکات دست و ترجمه آنها به دستورات حرکت پهپاد مورد استفاده قرار گرفته است. این روش امکان

کنترل دقیق و پویا برای پهباد را فراهم می‌کند و از قابلیت‌های پیشرفته یادگیری عمیق برای تشخیص حرکات دست بهره می‌برد.



شکل ۱-۲ چارچوب کنترل پهباد مبتنی بر ژست



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

نتیجه بدست آمده

این پروژه دقت بالایی در تشخیص ژست دست و کنترل پرواز پهباد دارد. پنج حالت دست مدنظر در این پروژه قرار گرفته‌اند و دقت متوسط آن برابر با ۹۷.۴۷۱ درصد است که نشان‌دهنده عملکرد بسیار عالی است. اما لازم به ذکر است که این دقت در پس‌زمینه‌های بهم ریخته و همچنین در شرایط نوری مختلف بسیار متغیر است زیرا ویژگی‌های Haar به سایه و رنگ‌های درون تصویر بسیار حساس است.

۲-۲-۲ مقاله A real-time hand gesture recognition method

در این مقاله در زمینه پردازش تصویر و تشخیص ژست‌های دست، از روش‌های مبتنی بر مدل ظاهری Appearance Model به عنوان یک رویکرد موثر استفاده شده است. این روش‌ها از ویژگی‌های تصویری

و حرکتی دست برای تشخیص و تعیین ژست‌های دست استفاده می‌کنند. در این مقاله، روش تشخیص ژست‌های دست به صورت زمان واقعی و قابل اعتماد است که از تشخیص دست، پیگیری دست^۳، تقسیم‌بندی دست^۴ و تشخیص ویژگی‌های مقیاس-فضا^۵ برای تشخیص ژست‌های دست استفاده می‌کند.

روش‌شناسی

در مقاله از ترکیب متنوعی از روش‌ها و ویژگی‌های تصویری استفاده می‌شود. این ویژگی‌ها به ترتیب به صورت زیر عمل می‌کنند.

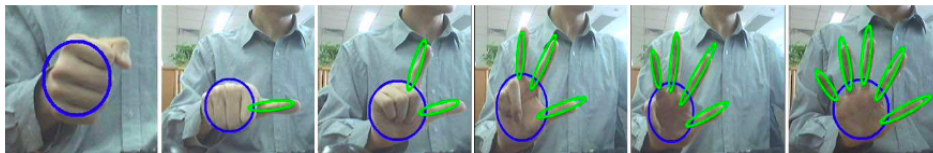
۱. استفاده از روش Adaboost برای تشخیص دست، که یک روش معتبر برای تشخیص اشیاء در تصاویر است.

۲. پیگیری دست با استفاده از تشخیص حرکت و رنگ، که از ترکیب تکنیک‌های جریان نوری و نشانه رنگ برای پیگیری دست در تصاویر استفاده می‌شود.

۳. تقسیم‌بندی دست با استفاده از اطلاعات حرکت و رنگ برای تمایز دست از پس‌زمینه و اشیاء دیگر.

۴. تشخیص ویژگی‌های مقیاس-فضا برای تشخیص ژست‌های دست، که برای شناسایی ساختارهای شبیه به کف دست و انگشتان استفاده می‌شود تا نوع ژست دست توسط ترکیب این ساختارها تعیین شود.

این روش‌ها و ویژگی‌ها با هم ترکیب شده‌اند تا یک سیستم تشخیص دست پایدار و دقیق برای استفاده در رابط کاربری تعاملی و تشخیص ژست‌های دست در زمینه‌های مختلف ارائه شود.



شکل ۲-۳ تشخیص کف دست و انگشتان

³Hand tracking

⁴Hand segmentation

⁵Scale-space feature detection

نتیجه

اعمال این روش‌ها نتایج قابل قبولی را به همراه داشته است. دقت مدل در تشخیص ژست‌های دست به صورت میانگین ۹۳.۸ درصد بوده و از جمله نتایج مهم آزمایشات می‌توان به تشخیص صحیح ۲۴۳۶ فریم از کل ۲۵۹۶ فریم ضبط شده اشاره کرد. این نتایج نشان می‌دهند که روش ارائه شده در این مقاله عملکرد قابل قبولی در تشخیص ژست‌های دست دارد و می‌تواند به عنوان یک روش موثر برای تعاملات زمان واقعی استفاده شود. [۱۴]

۳-۲-۲ مقاله Hand Gesture Recognition using Image Processing and Feature Extraction Techniques

روش‌شناسی

نتیجه

[۱۵]

۳-۲ مقالات مربوط به ورودی تصویر دست به مدل

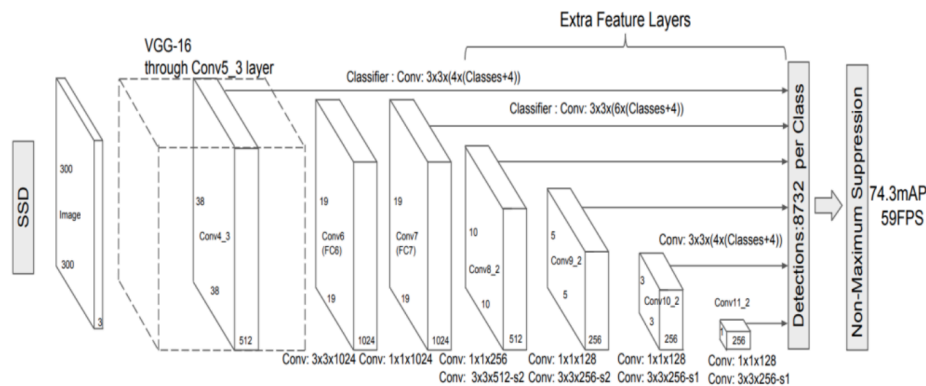
مقالات این دسته از جمله پروژه‌هایی هستند که بیشتر از آنکه بر روی پیش پردازش کار کنند باید بر روی معماری خود شبکه تمرکز کنند. در این مقالات تمام یا بخشی از تصویر گرفته شده به صورت یک ماتریس از تصویر با پیکسل‌های متعدد به مدل داده می‌شود. تنها موردی که می‌توان در این نوع پروژه‌ها پیش رو گرفت تشخیص موقعیت دست است تا بتوان تنها قسمتی از تصویر را به ورودی شبکه داد که دست در آن وجود دارد تا در حد ممکن اندازه ورودی شبکه و دقت آن افزایش یابد. پس از آن باید توجه داشت معماری شبکه را به گونه‌ای برگزید تا مخصوص پردازش تصویر باشد و بتوان ویژگی‌های تصویر را خود استخراج کند.

۱-۳-۲ مقاله Hand Gestures For Drone Control Using Deep Learning

این پروژه با هدف کنترل پهپادها با استفاده از حرکات دستی و با کمک شبکه‌های عمیق یادگیری انجام شده است تا بتوان ۹ حالت مختلف دست را شناسایی و به پهپاد، دستور موردنظر کاربر را داد.

روش‌شناسی

در این تحقیق، از معماری شبکه عصبی عمیق VGG-16 برای تشخیص و تعیین حرکات دستی برای کنترل پهپادها استفاده شده است. شبکه VGG-16 یکی از معروف‌ترین و پرکاربردترین شبکه‌های عمیق در زمینه بینایی ماشین است که شامل ۱۶ لایه عصبی با لایه‌های کانولوشنال و پولینگ می‌باشد. این شبکه برای استخراج ویژگی‌های مهم از تصاویر استفاده می‌شود. ورودی این شبکه تصاویری است که از دوربین متصل به دستگاه اجرایی گرفته می‌شود و سپس این تصاویر به شبکه وارد می‌شوند. خروجی این شبکه شامل تشخیص حرکات دستی مانند حرکات مختلف انگشتان و دست‌ها می‌شود که سپس این اطلاعات برای ارسال دستورات کنترلی به پهپاد استفاده می‌شود. این روش نه تنها امکان کنترل دقیق و موثر پهپادها را فراهم می‌کند بلکه ارتباط بین انسان و ماشین را نیز بهبود می‌بخشد.



شکل ۴-۲ معماری VGG-16

نتیجه بدست آمده

در پروژه این پروژه ۹ حالت دست مدنظر قرار گرفته‌شده و دقت آن برابر ۸۳.۳ درصد است. که در بهترین حالت ممکن با پس‌زمینه‌ی مناسب بدست آمده و باید در نظر گرفت که دقت بالایی برای کنترل پهپاد به حساب نمی‌آید. [۶]

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

Recognition

این مقاله به منظور کنترل پهپاد یا خلبان خودکار با استفاده از حرکت دست پیاده‌سازی شده است. به عنوان مثال، حرکت دست از چپ به راست نشان‌دهنده حرکت پهپاد به راست می‌باشد. برای اجرای این

برنامه، شبکه P-CNN طراحی شده است تا بتواند مفهوم تصاویر را تجربه کند.

روش‌شناسی

در این مقاله، از شبکه P-CNN^۶ برای تشخیص حرکات دست استفاده شده است. این شبکه از اطلاعات حرکت و ظاهر را بر روی مسیرهای بخش‌های بدن انسان (مانند دست راست، دست چپ، بدن بالا و بدن کامل) جمع‌آوری می‌کند. این شبکه ابتدا موقعیت دست فرد را با استفاده از جعبه مرزی مشخص می‌کند و سپس تصویر دست با استفاده از فیلترهای مناسب وارد شبکه P-CNN می‌شود تا بتواند حرکت دست را پیش‌بینی کند.

در خروجی این مدل، ۱۳ نوع حرکت مختلف وجود دارد که برای پیش‌بینی مفهوم آن‌ها استفاده می‌شود. این حرکات شامل کل دست از شانه تا انگشتان و حرکات آنها می‌شود، که در این پروژه برای دستور دادن به هواپیماهای بزرگ بدون سرنشین در فرودگاه‌ها استفاده می‌شود.

P-CNN اطلاعات ظاهر و حرکت را از بخش‌های مختلف بدن استخراج می‌کند و از دو شبکه پیش‌آموزش داده شده برای محاسبه ویژگی‌های CNN استفاده می‌کند. برای بخش‌های ظاهر، از شبکه VGG-f استفاده می‌شود، در حالی که برای بخش‌های حرکتی از شبکه حرکتی از پیاده‌سازی Action Tube استفاده می‌شود. ویژگی‌های استاتیک و پویا به طور جداگانه در طول زمان جمع‌آوری می‌شوند تا ویژگی‌های ویدیوی استاتیک و پویا به دست آید.

در نهایت، از روش‌های تجمیع مینیمم و ماکسیمم برای هر بعد از توصیف‌گر بر روی تمام ویدیوها استفاده می‌شود. این روش‌ها برای تشخیص حرکات با دقت بالا استفاده می‌شوند.

نتیجه

دارد، اما نیازمندی‌های پیچیده‌ای برای پیاده‌سازی واقعی دارد که از جمله موانعی است که باید در نظر گرفته شود. در نتیجه، این مقاله یک روش برای کنترل پهپاد با استفاده از حرکت دست پیاده‌سازی کرده است و از شبکه P-CNN برای تشخیص حرکات دست استفاده کرده است. نتایج نشان داده‌اند که این روش با دقت ۹۱.۹ درصد، قابلیت اجرا در پروژه‌های واقعی را دارد، اما نیازمندی‌های پیچیده‌ای برای پیاده‌سازی واقعی دارد که از جمله موانعی است که باید در نظر گرفته شود. [۱۶]

^۶Pose-based Convolutional Neural Network

۴-۲ مقالات مربوط به نقاط کلیدی دست

در این چنین مقالات ورودی شبکه بینایی ماشین برای تشخیص ژست، مختصات نقاط کلیدی دست هستند، که به نوعی یک ویژگی تصویر نیز تلقی می‌شوند. بدین صورت که در ابتدای کار دست کاربرد شناسایی شده و سپس نقاط کلیدی آن استخراج می‌شوند تا بتوان حجم داده ورودی به مدل را تا حد امکان ساده‌تر و در عین حال مفیدتر کرد.

۱-۴-۲ مقاله Hand Gesture Recognition system for Real-Time Application

[۱۷]

۲-۴-۲ مقاله An improved hand gesture recognition system using key-points and hand bounding boxes

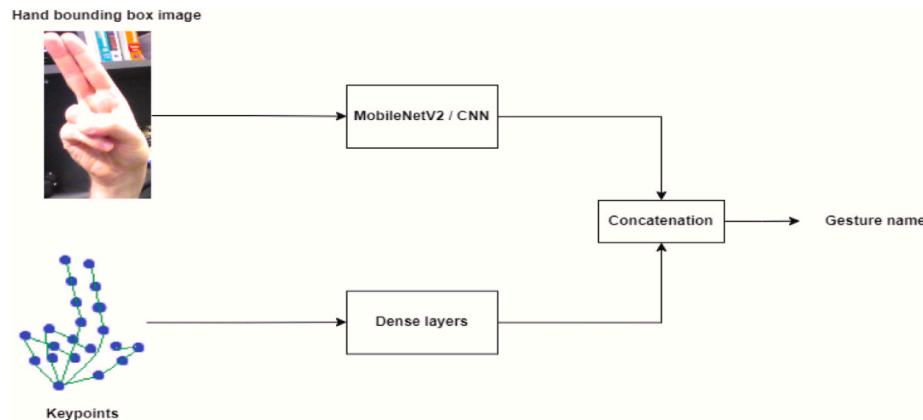
این مقاله یک سیستم بهبود یافته تشخیص حرکات دست با استفاده از نقاط کلیدی و جعبه‌های محدود کننده دست را معرفی می‌کند تا بتوان نقاط کلیدی دست را پیدا کرد.

روش‌شناسی

این پروژه از دو لوله موازی به نام‌های "MobileNetV2 + FC" و "CNN + FC" تشکیل شده‌است که تصاویر جعبه محدود کننده دست و ویژگی‌های استخراج شده از نقاط کلیدی را ترکیب می‌کند و از این طریق ژست دست را پیش‌بینی می‌کند.

در مدل MobileNetV2 + FC، از یک معماری سبک به نام MobileNetV2 برای استخراج ویژگی‌ها از تصاویر جعبه محدود کننده دست استفاده می‌شود. سپس این ویژگی‌ها به یک شبکه عصبی کاملاً متصل (FC) داده می‌شوند تا حرکات دست تشخیص داده شوند.

در مدل CNN + FC، در ابتدا نقاط کلیدی دست که اطلاعات مهمی درباره حرکات دست را شامل می‌شوند پیدا کرده و به یک شبکه عصبی کاملاً متصل (FC) وارد می‌شوند تا حرکات دست تشخیص داده شوند. این مدل از لایه‌های کانولوشنی برای کاهش تعداد پارامترها استفاده می‌کند و سپس از لایه‌های کاملاً متصل برای تشخیص حرکات دست استفاده می‌کند.



شکل ۵-۲ معماری ساختار شبکه‌های عصبی دو خط لوله

نتیجه

در صورتی که دو مدل خروجی‌های متفاوتی را پیش‌بینی کنند، از روش‌های ترکیبی مانند ترکیب احتمالاتی یا استفاده از مدل‌های متفاوت برای شرایط ورودی مختلف استفاده می‌شود تا بهترین تصمیم برای تشخیص ژست دست گرفته شود.

[۱۸]

۳-۴-۲ مقاله Visual gesture recognition based on hand key points

روش‌شناسی

نتیجه

[۱۹].

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

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 پرداخته شده 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 [۲۰]

۵-۲ مقالات مربوط به اجرای مدل‌های بینایی کامپیوتر روی

پهپاد

۱-۵-۲ مقاله Modeling relation among implementing AI-based drones and sustainable construction project success

روش‌شناسی

نتیجه

۶-۲ جمع‌بندی

[۲۱]

۲-۶-۱ مقاله

روش‌شناسی

نتیجه

[]

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

استخراج ویژگی‌های تصویر و یا ورودی خود تصویر به مدل می‌تواند بهینه عمل نکند

فصل سوم

روش انجام پروژه

۱-۳ مقدمه

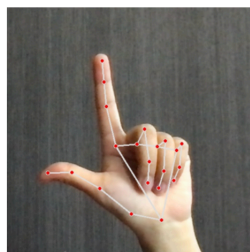
برای اجرای این پروژه راه‌های متفاوتی مورد بررسی قرار داده شد تا بتوان بهترین آنها را روی پهپاد پیاده‌سازی کرد. نتیجه نهایی پیاده‌سازی ۳ شبکه کانولوشن به صورت پی در پی است. شبکه اول برای آشکارسازی موقعیت کف دست است. بدین صورت که هر فریم گرفته‌شده از دوربین پهپاد پس از تغییر اندازه به یک ماتریس $256 \times 256 \times 3$ به ورودی مدل داده می‌شود و پس از پردازش آن خروجی یک ماتریس $256 \times 256 \times 3$ است که جعبه محدودکننده دست را شامل می‌شود. پس از آن، این ماتریس به مدل دوم به عنوان ورودی داده می‌شود، در این مدل جعبه مرزی برش خورده 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 $S =$ 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 انتخاب ژست‌های دست متناسب با حرکت پهپاد

انتخاب ژست‌های مناسب برای هر یک از حرکات پهپاد از اهمیت ویژه‌ای برخوردار است. چرا که ژست‌هایی که از لحاظ مفهومی به عملکرد پهپاد شبیه هستند راحت‌تر به خاطر سپرده شده و تجربه دلپذیرتری را در کاربر به وجود می‌آورند. ما برای این پروژه ۹ ژست دست را در نظر گرفته‌ایم تا بتوان حرکات پایه پهپاد را با آنها انجام داد. این حرکات شامل: حرکت رو به جلو، حرکت رو به عقب، حرکت به پایین، حرکت به بالا، حرکت به راست، حرکت به چپ، فرود آمدن، ایستادن در موقعیت کنونی و گرفتن عکس است که نمونه ژست دست آنها با توجه به حرکت پهپاد نشان داده شده است.

۳-۳ دیتاست

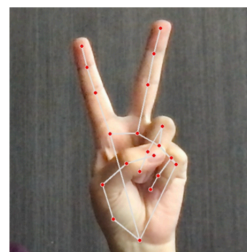
برای جمع‌آوری دیتاست مناسب پروژه، از آنجایی که ژست‌های دست را بر اساس عملکرد پهپاد تعیین کردیم تا استفاده از آنها برای کاربر مورد پسند باشند، لذا پیدا کردن دیتاست آماده ناممکن است.



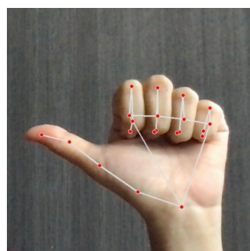
بالا 🖐️



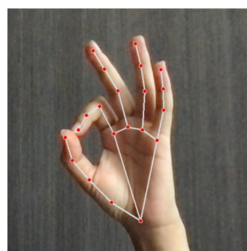
پایین 🖐️



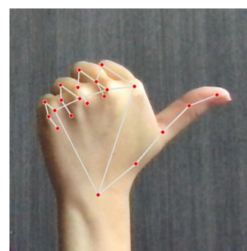
گرفتن عکس 📷



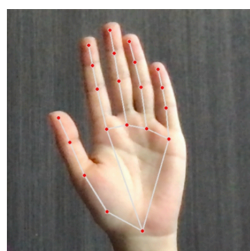
چپ 🖐️



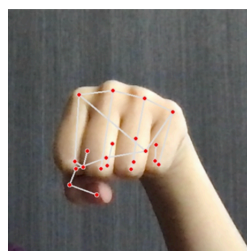
فرود 🖐️



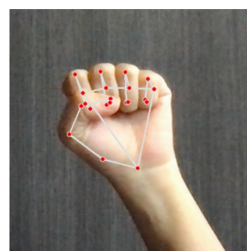
راست 🖐️



جلو 🖐️



عقب 🖐️



ایست 🖐️

شکل ۳-۱ نمونه‌ای از ژست‌های انتخاب‌شده در مجموعه داده‌ها

جمع‌آوری دیتاست مناسب و جامع از اهمیت بالایی برخوردار است زیرا از آنجایی که مدل باید در زمان واقعی کار کند نمی‌توان از لایه‌های زیادی استفاده کرد. پس باید برای بالا بردن دقت مدل از موارد دیگری کمک گرفت، از جمله آنها می‌توان به پیش‌پردازش، پس‌پردازش و دیتاست مناسب اشاره کرد. از آنجایی که رویکردهای متفاوتی برای این پروژه پیاده‌سازی شد تا بتوان به بهترین راهکار برای مدلی سبک و در عین حال جامع رسید، دیتاست‌های گوناگونی جمع‌آوری شد که هر یک از آنها ویژگی خاصی از تصویر را به عنوان دیتای مرود نیاز جمع‌آوری می‌کردند. از جمله ورودی کل عکس به صورت پیکسل‌های رنگی، پیدا کردن دست و ذخیره موقعیت آن به صورت پیکسل‌های 256×256 ، پیدا کردن نقاط کلیدی دست و ذخیره موقعیت آنها در دیتاست و استفاده از آنها برای ورودی به مدل بهترین گزینه ممکن بود مه برای اجرای پروژه از آن استفاده کردیم.

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^۱ 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 کنیم، فکر کنیم، و به خاطر بسپاریم، گاهی دانش منحصر به فردی را که هنوز نمی توان به زبان آورد، آشکار می کنند. به طوری که می توان گفت ژست ها اغلب به عنوان زبان گفتاری ثانویه در نظر گرفته می شود. [۲۲] ژست ها به ویژه زمانی مؤثر هستند که مزیتی نسبت به کلمات داشته باشند. [۲۳] توانایی درک شکل و حرکت دست ها می تواند یک جزء حیاتی در بهبود تجربه کاربر^۲ در حوزه ها و پلتفرم های 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 دست ها اغلب خود یا یکدیگر را مسدود می کنند مانند انسداد انگشت، کف دست و لرزش دست و فاقد الگوهای کنتراست بالا هستند. [۲۰]

۵-۳ کنترل پهباد

اکثر پهبادهای تجاری موجود در بازار یا دارای کنترلرهای طراحی شده ویژه هستند، یا دارای فرستنده سیگنال اختصاصی و برنامه های نرم افزاری هستند که روی دستگاه های دستی کاربران مانند تلفن های همراه یا تبلت ها اجرا می شوند. در هر دو مورد، کنترل کننده فرمان هایی را با اطلاعات دقیق از طریق کانال های بی سیم مانند وای فای یا بلوتوث ارسال می کند. اخیراً محصولات تجاری وجود داشته است که حرکات دست را به عنوان یک مکانیسم کنترل قابل اجرا معرفی می کنند. برای گرفتن ژست ها، دو رویکرد وجود دارد.

• استفاده از دستکش های طراحی شده ویژه: کنترل کننده بر روی دستکشی که توسط کاربران

¹co-speech gestures

²user experience

استفاده می شود نصب می شود و در زمان واقعی انحراف، گام و چرخش دست را شناسایی می کند 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

۳-۶ ابزارها و نرم افزارهای 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 برای برنامه نویسان یادگیری ماشین از جمله محققان، دانشجویان و توسعه دهندگان نرم افزار، که برنامه های کاربردی یادگیری ماشین را پیاده سازی می کنند، نمونه های اولیه فناوری را طراحی می کند تا بتوان پروژه ها را تا حد امکان ساده کرد. برنامه هایی که داده های حسی مثل ویدیو و صدا را با نرخ فریم بالا پردازش می کند تا تجربه کاربر را بهتر کند. مراحل پردازش یا مدل های استنتاجی ممکن است دشوار باشد، چون گاهی اتصال بین مراحل زیاد است. همچنین، توسعه برنامه برای پلتفرم زمان بر است. [۲۴]

Media Pipe این چالش ها را با انتزاع و اتصال مدل های مختلف به یکدیگر در یک چارچوب مناسب حل می کند. با استفاده از MediaPipe، می توان یک لوله پردازش را به صورت گراف از اجزای مختلف، از جمله مدل های استنتاجی و عملکردهای پردازش رسانه ای، ساخت. همچنین این کتابخانه می تواند مطابق با نیازهای افراد خود سفارشی شود و در پلتفرم های مختلف توسعه پیدا کند [۲۵].

در مجموعه MediaPipe نیز از کتابخانه های مختلفی برای پیاده سازی برنامه ها استفاده می شود. از جمله

آنها می‌توان به TensorFlow، PyTorch، OpenCV، CNTK و MXNet اشاره کرد. [۲۵]

۷-۳ مدیاپایپ

برای پیاده سازی شبکه‌های تشخیص کف دست و پیدا کردن نقاط عطف دست از مدل‌های از قبل آموزش دیده^۳ کتابخانه مدیاپایپ کمک گرفته شده است. مدیاپایپ از یک خط لوله یادگیری ماشین متشکل از چندین مدل که با هم کار می‌کنند استفاده می‌کند: یک مدل تشخیص کف دست^۴ که تصویر را از ورودی می‌گیرد و عکس محدوده دست را به عنوان خروجی دریافت میکند و یک مدل تشخیص نقاط عطف دست^۵ که عکس دست را به عنوان ورودی گرفته و مختصات ۲۱ نقطه کلیدی بندهای انگشتان دست را در ناحیه دست تشخیص می‌دهد.

۱-۷-۳ مدل تشخیص کف دست

مدل تشخیص کف دست مدیاپایپ دارای دقت متوسط ۹۵.۷ درصد است که این دقت بالا با استفاده از استراتژی‌های مختلف به دست آمده است. ابتدا، به جای آشکار کردن دست^۶، 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^۷ که یک تکنیک پس پردازش^۸ است و در تشخیص اشیاء برای حذف تشخیص‌های تکراری^۹ و انتخاب مرتبط ترین اشیاء شناسایی شده استفاده می‌شود. این به کاهش مثبت کاذب^{۱۰} و پیچیدگی محاسباتی^{۱۱} یک الگوریتم تشخیص کمک می‌کند. تا بهترین محدوده مربعی^{۱۲} با واریانس بالا^{۱۳} را بدست آورد. [۲۰]

³Pretrained

⁴Palm detection model

⁵Hand landmark model

⁶hand detector

⁷Non-maximum suppression

⁸post-process

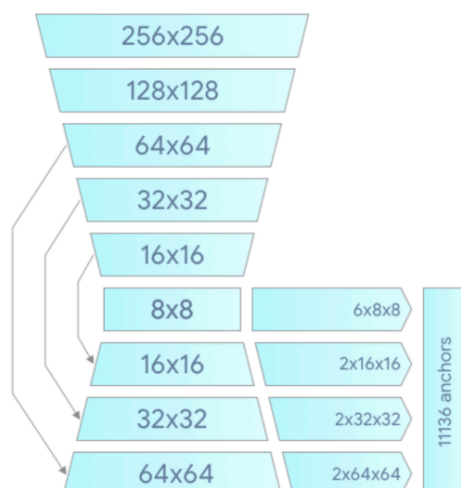
⁹duplicate detections

¹⁰false positive

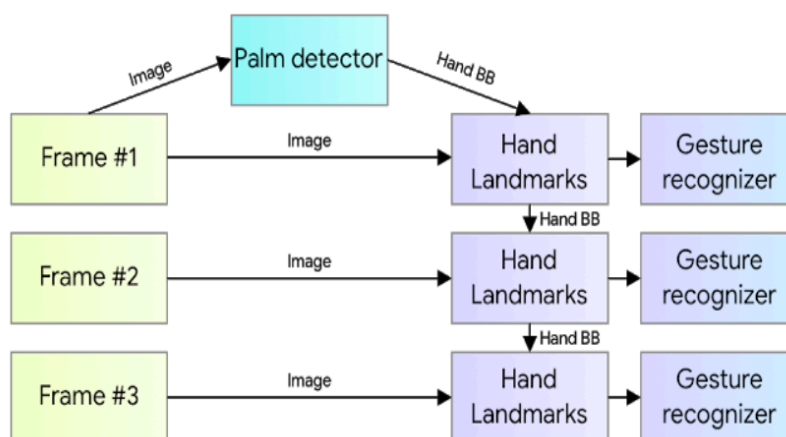
¹¹computational complexity

¹²bounding box

¹³high scale variance



شکل ۲-۳ معماری مدل آشکارساز کف دست



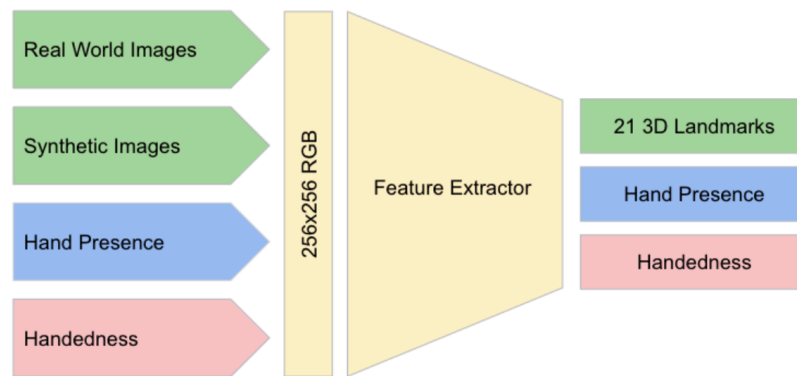
شکل ۳-۳ خط لوله تشخیص دست

۲-۷-۳ مدل تشخیص نقاط عطف دست

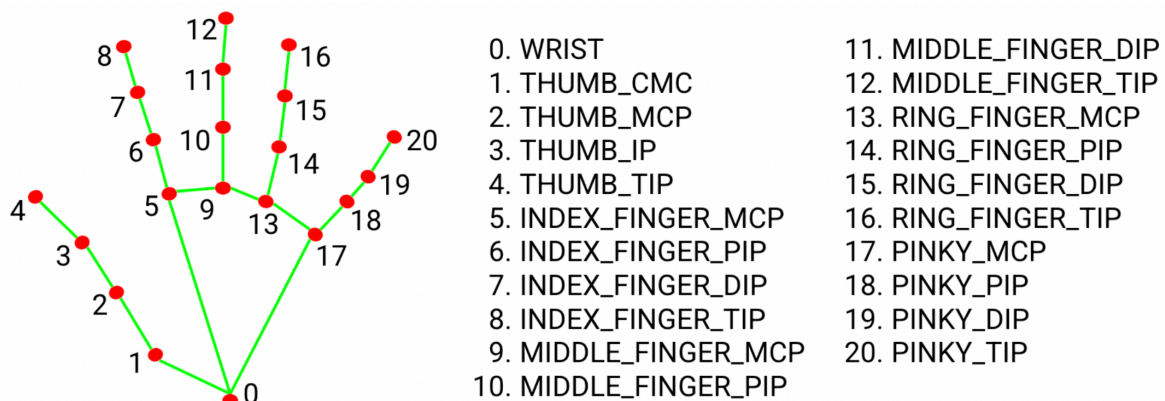
در این مرحله مکان‌یابی مختصات ۲۱ نقطه کلیدی بندهای انگشتان دست که شامل سه بعد است از طریق رگرسیون^{۱۴} انجام می‌شود. این مدل بر روی ۳۰ هزار تصویر دنیای واقعی با ۲۱ مختصات سه بعدی برچسب زده‌شده^{۱۵} آموزش دیده‌است. برای پوشش بهتر ژست‌های احتمالی دست و ارائه نظارت بیشتر بر ماهیت هندسه دست، این دیتاست از مدل‌های دست مصنوعی با کیفیت بالا را نیز روی پس‌زمینه‌های مختلف ارائه می‌کند تا دقت را به بالاترین حد ممکن برساند. این مدل حتی در برابر دست‌های نیمه نیز عملکرد قوی نشان می‌دهد. [۲۰]

¹⁴regression

¹⁵labeling



شکل ۳-۴ 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 ۳-۵ موقعیت ۲۱ نقطه کلیدی در ناحیه دست

۸-۳ پیش پردازش

برای ورود نقاط عطف دست به مدل تعیین ژست به ۲۱ مختصات \square و \square نیاز داریم. خروجی مدل تعیین مختصات نقاط عطف دست برابر مختصات مطلق پیکسل ها نسبت به گوشه سمت چپ پایین تصویر است. این نقاط با توجه به اندازه عکس می توانند گسترده باشند برای مثال در یک عکس با اندازه 2048×2048 این اعداد از بین ۰ تا ۲۰۴۸ متغیر اند. اگر این مختصات را به صورت مستقیم به مدل تعیین ژست دست بدهیم دقت مدل برابر ۸۷ درصد خواهد بود که به میزان کافی مورد قبول نیست. برای بهبود آن باید پیش پردازش هایی بر روی داده ورودی انجام شود.

از جمله این پیش پردازش ها می توان به نسبی کردن و نرمال سازی داده ها اشاره کرد. برای این کار ابتدا باید یک مرجع واحد در نظر گرفت تا نقاط، نسبت به آن مشخص شوند. در این پروژه ما مرجع را نقطه مشخص شده روی مچ در نظر می گیریم. مختصات نقطه مرجع را برابر $(0,0)$ قرار می دهیم. سپس نسبت

به آن و با توجه به فرمول زیر مختصات نقاط دیگر را به روز رسانی می‌کنیم.

$$X_{rel} = X_{ref} - X$$

پس از به نسبی کردن نقاط نسبت به مبدأ، آنها را با کمک فرمول زیر نرمال سازی می‌کنیم تا تمام طول و عرض نقاط به عددی میان صفر و یک به روز رسانی شوند.

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

در انتها این مختصات را به عنوان ورودی به شبکه تعیین ژست دست می‌دهیم. با توجه به اینکه معماری هیچ یک از مدل‌ها تغییر نکرد و تنها داده‌های مختصات به روز رسانی شدند، دقت نهایی مدل به ۹۷ درصد افزایش پیدا کرد و پیش‌پردازش تاثیر به‌سزایی در بهینه کردن پروژه داشت.

۹-۳ رأی‌گیری پنجره‌ای

با وجود اینکه دقت مدل پیاده‌سازی شده بالا است و عملکرد بسیار چشم‌گیری از خود نشان می‌دهد، در عین حال پیش‌بینی اشتباه مدل می‌تواند عواقب زیان‌باری را به ارمغان آورد، از تجربه ناپسند برای کاربر گرفته تا برخورد پهلپاد به اجسام و هزینه مالی. لذا باید دقت انجام پروژه را از آنچه مدل پیش‌بینی می‌کند نیز بالاتر برد. برای این کار از رأی‌گیری پنجره‌ای استفاده کرده‌ایم. بدین صورت که متغیری را با توجه به FPS دوربین در نظر می‌گیریم بدین صورت که هر چه FPS دوربین پهلپاد بیشتر باشد متغیر در نظر گرفته‌شده نیز بیشتر است، برای مثال در پروژه ما از آنجایی که دوربین پهلپاد برابر ۳۰ FPS است ما این متغیر را ۱۰ قرار داده‌ایم. سپس حد مناسبی را نیز بین صفر تا یک قرار می‌دهیم که ما در پروژه آن را برابر ۷۰۰ قرار داده‌ایم. طبق این راه ما ۱۰ فریم متناوب گرفته‌شده از پهلپاد را به مدل پیاده‌سازی شده می‌دهیم اما تنها در صورتی دستور پیش‌بینی شده را به پهلپاد پهلپاد می‌دهیم که حد ممکن را بدست آورند. برای مثال اگر ۷ یا بیشتر از ۱۰ حرکت پیش‌بینی شده، دستور حرکت رو به جلو باشد آنگاه به پهلپاد دستور داده می‌شود تا به جلو حرکت کند. در غیر این صورت اگر کمتر از ۷ عدد از فریم‌ها یک ژست دست را پیش‌بینی نکنند، پهلپاد در حالت قبلی خود باقی می‌ماند و دستوری به آن داده نمی‌شود.

فصل چهارم

نتایج و ارزیابی

۱-۴ مقدمه

این پروژه با مدل‌های گوناگونی پیاده‌سازی شد تا بتوان بهترین آنها را برای نتیجه نهایی بر روی پهنپاد اجرا کرد. نمونه خروجی این مدل‌ها به شرح زیر است.

معماری مدل	دوره ^۱	دقت آموزش	دقت تست
MLP	۶	۸۷۸۳۷	۷۸۷
CNN	۷	۷۸	۵۴۱۵
LSTM	۵۴۵	۷۷۸	۷۵۰۷
۴	۵۴۵	۱۸۷۴۴	۷۵۶۰
۵	۸۸	۷۸۸	۶۳۴۴

جدول ۱-۴ جدول

۲-۴ نتایج و ارزیابی

۳-۴ جمع‌بندی

فصل پنجم

نتیجه گیری و پیشنهادات

۱-۵ مقدمه

۲-۵ نتیجه گیری و پیشنهادات

۳-۵ جمع بندی

کتاب نامه

- [1] Walter, Ian and Khadr, Monette. Gesture controlled drone.
- [2] Puri, Vikram, Nayyar, Anand, and Raja, Linesh. Agriculture drones: A modern breakthrough in precision agriculture. *Journal of Statistics and Management Systems*, 20(4):507–518, 2017.
- [3] Gatteschi, Valentina, Lamberti, Fabrizio, Paravati, Gianluca, Sanna, Andrea, Demartini, Claudio, Lisanti, Alberto, and Venezia, Giorgio. New frontiers of delivery services using drones: A prototype system exploiting a quadcopter for autonomous drug shipments. in *2015 IEEE 39th annual computer software and applications conference*, vol. 2, pp. 920–927. IEEE, 2015.
- [4] Moore, T. Nypd considering using drones to fight crime. *New York Daily*, 2014.
- [5] Natarajan, Kathiravan, Nguyen, Truong-Huy D, and Mete, Mutlu. Hand gesture controlled drones: An open source library. in *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, pp. 168–175. IEEE, 2018.
- [6] Hadri, Soubhi. Hand gestures for drone control using deep learning. 2018.
- [7] Zhu, Pengfei, Wen, Longyin, Bian, Xiao, Ling, Haibin, and Hu, Qinghua. Vision meets drones: A challenge. *arXiv preprint arXiv:1804.07437*, 2018.
- [8] Guvenc, Ismail, Koohifar, Farshad, Singh, Simran, Sichitiu, Mihail L, and Matolak, David. Detection, tracking, and interdiction for amateur drones. *IEEE Communications*

- Magazine*, 56(4):75–81, 2018.
- [9] Hassanalian, Mostafa and Abdelkefi, Abdessattar. Classifications, applications, and design challenges of drones: A review. *Progress in Aerospace sciences*, 91:99–131, 2017.
- [10] Yoo, Minjeong, Na, Yuseung, Song, Hamin, Kim, Gamin, Yun, Junseong, Kim, Sangho, Moon, Changjoo, and Jo, Kichun. Motion estimation and hand gesture recognition-based human–uav interaction approach in real time. *Sensors*, 22(7):2513, 2022.
- [11] Ma, Yuntao, Liu, Yuxuan, Jin, Ruiyang, Yuan, Xingyang, Sekha, Raza, Wilson, Samuel, and Vaidyanathan, Ravi. Hand gesture recognition with convolutional neural networks for the multimodal uav control. in *2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*, pp. 198–203. IEEE, 2017.
- [12] Hu, Bin and Wang, Jiacun. Deep learning based hand gesture recognition and uav flight controls. *International Journal of Automation and Computing*, 17(1):17–29, 2020.
- [13] Sarkar, Ayanava, Patel, Ketul Arvindbhai, Ram, RK Ganesh, and Capoor, Geet Krishna. Gesture control of drone using a motion controller. in *2016 international conference on industrial informatics and computer systems (ciics)*, pp. 1–5. IEEE, 2016.
- [14] Fang, Yikai, Wang, Kongqiao, Cheng, Jian, and Lu, Hanqing. A real-time hand gesture recognition method. in *2007 IEEE International Conference on Multimedia and Expo*, pp. 995–998. IEEE, 2007.
- [15] Sharma, Ashish, Mittal, Anmol, Singh, Savitoj, and Awatramani, Vasudev. Hand gesture recognition using image processing and feature extraction techniques. *Procedia Computer Science*, 173:181–190, 2020.
- [16] Perera, Asanka G, Wei Law, Yee, and Chahl, Javaan. Uav-gesture: A dataset for uav

- control and gesture recognition. in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0, 2018.
- [17] Murugeswari, M and Veluchamy, S. Hand gesture recognition system for real-time application. in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, pp. 1220–1225. IEEE, 2014.
- [18] Dang, Tuan Linh, Tran, Sy Dat, Nguyen, Thuy Hang, Kim, Suntae, and Monet, Nicolas. An improved hand gesture recognition system using keypoints and hand bounding boxes. *Array*, 16:100251, 2022.
- [19] Chen, Boxu, Yu, Lixin, Meng, Xiao, and Hua, Yang. Visual gesture recognition based on hand key points. in *Journal of Physics: Conference Series*, vol. 2024, p. 012037. IOP Publishing, 2021.
- [20] Zhang, Fan, Bazarevsky, Valentin, Vakunov, Andrey, Tkachenka, Andrei, Sung, George, Chang, Chuo-Ling, and Grundmann, Matthias. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020.
- [21] Waqar, Ahsan, Othman, Idris, Hamah Sor, Nadhim, Alshehri, Abdullah Mohammed, Almujiabah, Hamad R, Alotaibi, Badr Saad, Abuhussain, Mohammed Awad, Bageis, Abdulrahman S, Althoey, Fadi, Hayat, Saleh, et al. Modeling relation among implementing ai-based drones and sustainable construction project success. *Frontiers in Built Environment*, 9:1208807, 2023.
- [22] Clough, Sharice and Duff, Melissa C. The role of gesture in communication and cognition: Implications for understanding and treating neurogenic communication disorders. *Frontiers in Human Neuroscience*, 14:323, 2020.
- [23] Kang, Seokmin and Tversky, Barbara. From hands to minds: Gestures promote understanding. *Cognitive Research: Principles and Implications*, 1:1–15, 2016.

- [24] Lugaresi, Camillo, Tang, Jiuqiang, Nash, Hadon, McClanahan, Chris, Uboweja, Esha, Hays, Michael, Zhang, Fan, Chang, Chuo-Ling, Yong, Ming, Lee, Juhyun, et al. Mediapipe: A framework for perceiving and processing reality. in *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, vol. 2019, 2019.
- [25] Harris, Moh, Agoes, Ali Suryaperdana, et al. Applying hand gesture recognition for user guide application using mediapipe. in *2nd International Seminar of Science and Applied Technology (ISSAT 2021)*, pp. 101–108. Atlantis Press, 2021.