

تشخیص مکان و زاویه گل‌های زعفران در تصویر با استفاده از شبکه‌های همگشتی عمیق

محمدرضا محمدی^۱ و مهرداد کرمی^۲

چکیده

تشخیص خودکار مکان و زاویه برش گل‌های زعفران یک گام اساسی برای فراآوری خودکار گل زعفران است. با توجه به تنوع موجود میان گل‌های زعفران، الگوریتم‌های سنتی نمی‌توانند دقت قابل قبول را نتیجه دهند. در این مقاله، یک الگوریتم مبتنی بر شبکه‌های همگشتی عمیق برای حل این مسئله پیشنهاد شده است. در معماری پیشنهادی، تصویر به تعدادی ناحیه مربعی کوچک تقسیم می‌شود و در هر ناحیه وجود محل برش گل زعفران تخمین زده می‌شود. همچنین، محل نسبی مناسب برای برش گل زعفران نسبت به مرکز ناحیه تخمین زده می‌شود. برای تخمین زاویه، ابتدا محدوده ۳۶۰ درجه به تعدادی قسمت تقسیم می‌شود و زاویه گل به یکی از مقادیر اختصاص می‌یابد. سپس، زاویه نسبی نسبت به مرکز ناحیه تخمین زده می‌شود. به منظور ارزیابی عملکرد الگوریتم پیشنهادی، یک پایگاه داده شامل ۱۶۳ تصویر و ۳۰۳۵ گل جمع‌آوری شده است و پارامترهای هر گل توسط یک خبره مشخص شده است. ارزیابی الگوریتم پیشنهادی نشان می‌دهد که پارامترهای بیش از ۹۵٪ گل‌ها به درستی تخمین زده شده‌اند.

کلیدواژه‌ها

گل زعفران، یادگیری عمیق، شبکه‌های همگشتی عمیق، تشخیص هدف

فرآیند می‌تواند قیمت تمام شده محصول را کاهش دهد و منجر به رونق بیشتر صنعت کشت زعفران در کشور شود.

مطالعه برای خودکارسازی فرآیند فراآوری زعفران با استفاده از تکنیک‌های پردازش تصویر از سال‌ها قبل شروع شده است. در [۲] ابتدا با استفاده از یک آستانه‌گذاری ساده، ناحیه مربوط به گل زعفران از زمینه جدا شده است. سپس، با تحلیل شعاع دایره‌های محاطی در مرز تصویر باینری، محل مناسب برای برش تشخیص داده شده است. به طور مشابه، در [۳] نیز ابتدا تصویر دوسطحی می‌شود و سپس با استفاده از تحلیل اسکلت تصویر، محل مناسب برای برش گل زعفران تخمین زده می‌شود. در [۴] نیز با استفاده از تکنیک‌های سنتی پردازش تصویر، ابتدا ناحیه مربوط به گل زعفران در تصویر ورودی مشخص می‌شود. سپس، از ناحیه تشخیص داده شده تعدادی ویژگی استخراج می‌شود و برای آموزش یک شبکه عصبی با ۱ لایه مخفی استفاده می‌شود.

۱ مقدمه

یکی از فرآیندهای مهم در فراآوری زعفران، جداکردن قسمت ارزشمند گیاه، یعنی کلاله، از سایر اجزای گل است که این موضوع در شکل ۱ نشان داده شده است. بر اساس آمارهای بین‌المللی، حدود ۹۰ درصد از تولید زعفران جهان در ایران انجام می‌شود [۱]. در حال حاضر، فرآیند فراآوری زعفران به صورت دستی انجام می‌شود که دارای چالش‌های بسیار زیادی است. خودکارسازی این

این مقاله در مردادماه سال ۱۳۹۹ دریافت، در مهرماه بازنگری و در آبان‌ماه پذیرفته شد.

^۱ دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

رایانامه: mrmohammadi@iust.ac.ir

^۲ شرکت سامانه‌های هوشمند کشاورزی والی

رایانامه: karami@valisystem.ir

نویسنده مسئول: محمدرضا محمدی

DOR: 20.1001.1.23831197.1400.8.3.4.2

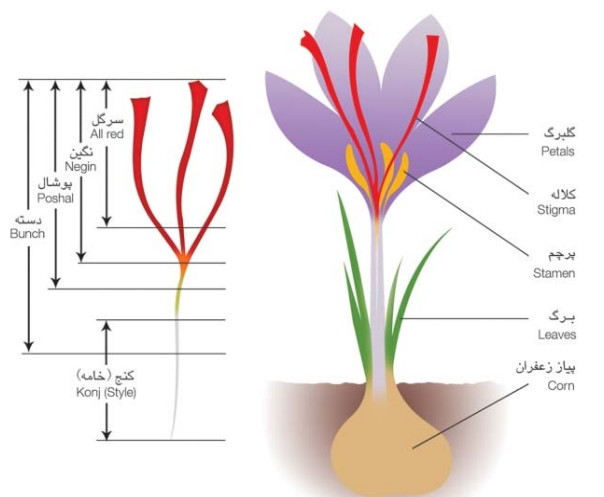
- تهیه یک پایگاه داده دارای حاشیه‌نویسی از تصاویر گل‌های زعفران
- توسعه یک معماری مبتنی بر شبکه‌های همگشتی عمیق برای تخمین خودکار مکان و زاویه برش گل‌های زعفران در سامانه طراحی شده به منظور خودکارسازی فرآیند پاک‌سازی گل‌های زعفران، ابتدا گل‌های زعفران بر روی یک نوار نقاله قرار می‌گیرند. سپس، از نوار نقاله تصویربرداری می‌شود و با استفاده از الگوریتم پیشنهادی در این مقاله، مکان و زاویه گل‌ها تشخیص داده می‌شوند. در نهایت، توسط چندین بازوی مکانیکی، گل‌های زعفران برش می‌خورند تا قسمت کلاله از سایر اجزای گل جدا شود و فرآیند پاک‌سازی گل زعفران تکمیل شود. تمرکز این مقاله منحصر به بخش بینایی کامپیوتر برای تشخیص مکان و زاویه گل‌های زعفران است.

ادامه مقاله به صورت زیر تدوین شده است. در بخش ۲ مهمترین روش‌های موجود برای آشکارسازی اشیاء مرور می‌شوند. روش پیشنهادی برای تخمین مکان و زاویه برش گل‌های زعفران در بخش ۳ ارائه می‌شود. پایگاه داده جمع‌آوری شده و نتایج استفاده از الگوریتم پیشنهادی در بخش ۴ گزارش می‌شود و جمع‌بندی مقاله در بخش ۵ انجام می‌شود.

۲ مرور کارهای پیشین

شبکه‌های همگشتی عمیق در ابتدا برای کاربرد دسته‌بندی تصویر توسعه یافتند. در ادامه و با توجه به نتایج مناسبی که توسط این شبکه‌ها بدست آمد، از این ابزار در کاربردهای دیگر حوزه بینایی کامپیوتر نیز استفاده شده است. تشخیص مکان و زاویه مناسب برای برش گل‌های زعفران با استفاده از شبکه‌های همگشتی عمیق تا کنون مورد مطالعه قرار نگرفته است. با این وجود، حوزه آشکارسازی اشیاء مشابه‌ترین حوزه با مسئله مورد مطالعه در این مقاله است که روند پیشرفت آن در ادامه بررسی می‌شود.

الگوریتم R-CNN در [۹] پیشنهاد شده است که در آن ابتدا با استفاده از روش SelectiveSearch [۱۹] تعدادی ناحیه پیشنهادی از تصویر ورودی استخراج می‌شود و سپس برچسب هر ناحیه پیشنهادی توسط یک CNN پیش‌بینی می‌شود. به عبارت دیگر، این روش شامل دو بخش است که بخش اول به صورت مستقل ناحیه‌های پیشنهادی را استخراج می‌کند و بخش دوم عمل دسته‌بندی تصویر را انجام می‌دهد. با وجود دقت قابل توجه، به دلیل فراخوانی متعدد یک شبکه عمیق در این رویکرد، هزینه محاسباتی این الگوریتم بسیار زیاد است و برای بسیاری از کاربردهای زمان حقیقی قابل استفاده نیست. به منظور افزایش سرعت، الگوریتم Fast R-CNN در [۲۰] پیشنهاد شد تا محاسبات مربوط به لایه‌های همگشتی تنها یک بار برای تمام تصویر انجام شود و سپس ویژگی‌های مربوط به هر ناحیه پیشنهادی از آن استخراج شده و به دسته‌بند وارد شود. سپس، با اضافه کردن بخش آموزش تولید ناحیه‌های پیشنهادی، الگوریتم



شکل ۱ اجزای گل زعفران [۵]

یک ایراد اساسی در روش‌های ذکر شده این است که فرض می‌کنند گل‌های زعفران دارای همپوشانی نیستند. همچنین، تصاویر بررسی شده دارای زمینه ساده‌ای هستند که دوسطحی کردن تصویر را با مشکل جدی روبرو نمی‌کند. با این وجود، در کاربردهای واقعی همپوشانی گل‌ها اجتناب‌ناپذیر است و همچنین، به دلیل برش‌های متعددی که انجام می‌شود، تصویر زمینه ثابت نخواهد ماند که جداسازی مرز دقیق گل زعفران را با چالش روبرو می‌کند. به منظور حل این محدودیت‌ها، در این مقاله از روش یادگیری عمیق استفاده می‌شود.

الگوریتم‌های مبتنی بر یادگیری عمیق در حوزه‌های مختلف بینایی کامپیوتر از جمله دسته‌بندی تصویر [۶-۸]، آشکارسازی اشیاء [۹-۱۰]، ناحیه‌بندی معنایی [۱۱-۱۲] و مکان‌یابی نقاط مشخصه [۱۳-۱۵] بهترین نتایج را در سال‌های اخیر بدست آورده‌اند. ایده اصلی در موفقیت شبکه‌های عمیق، یادگیری سلسله‌مراتبی مسئله است. البته به طور ویژه در حوزه بینایی کامپیوتر، استفاده از لایه‌های همگشتی است که با در نظر گرفتن وابستگی مکانی پیکسل‌ها و به اشتراک گذاشتن وزن‌ها، موفقیت‌های بسیاری را به ارمغان آورده است.

شبکه‌های همگشتی عمیق در حوزه کشاورزی و صنایع غذایی نیز دارای کاربردهای متنوعی است که در سال‌های اخیر مورد توجه بیشتری قرار گرفته است. در [۱۶] از یک شبکه همگشتی عمیق برای دسته‌بندی نوع میوه از میان ۶۰ کلاس استفاده شده است. به منظور تشخیص موقعیت میوه‌ها بر روی درخت، در [۱۷-۱۸] از دو شبکه همگشتی عمیق استفاده شده است (به ترتیب دو و سه نوع میوه متفاوت مورد بررسی قرار گرفته است). در این دو مطالعه، هدف آشکارسازی مستطیل‌هایی در تصویر است که شامل میوه‌های مورد نظر هستند.

در این مقاله، هدف بدست آوردن مکان و زاویه برش برای گل‌های زعفران موجود در تصویر ورودی است که در مطالعات قبلی چنین الگوریتمی مشاهده نشده است. دستاوردهای اصلی مقاله عبارتند از:

بنابراین، خروجی نهایی شبکه دارای ابعاد $4 \times H/2^b \times W/2^b$ خواهد بود. برای تخمین این چهار پارامتر نیز از چند لایه همگشتی استفاده می‌شود. استفاده از لایه‌های همگشتی بیشتر منجر به افزایش ناحیه موثر^۱ و افزایش تعداد پارامترهای شبکه خواهد شد. ما در پیاده‌سازی‌های خود از ۵ لایه همگشتی برای این بخش استفاده کرده‌ایم.

به منظور آموزش شبکه، نیاز است تا برای هر تصویر آموزشی، خروجی مطلوب شبکه تعیین شود. در پایگاه داده مورد استفاده، مکان و زاویه تمام گل‌های زعفران ثبت شده‌اند. برای هر کدام از خروجی‌هایی که فاصله مرکز مربوطه نسبت به یکی از گل‌ها از حدی کمتر باشد، احتمال ۱ و برای باقی خروجی‌ها احتمال ۰ در نظر گرفته می‌شود. سپس، اختلاف مکان گل نسبت به مرکز انتخاب شده به عنوان خروجی‌های مطلوب مختصات لحاظ می‌شوند و زاویه گل به عنوان پارامتر چهارم تعیین می‌شود. پس از مشخص شدن خروجی مطلوب، تابع ضرر از مقایسه آن با خروجی شبکه محاسبه می‌شود:

$$L = L_{prob} + L_{loc} + L_{ang} \quad (۱)$$

که L_{prob} تابع ضرر مربوط به تخمین احتمال وجود نقطه برش گل است. با توجه به اینکه درصد نقاطی که شامل نقطه برش هستند عدد کوچکی است، نمونه‌های آموزشی دارای برچسب ۱ به مراتب کمتر از نمونه‌های آموزشی دارای برچسب ۰ است. به همین دلیل، در این مقاله از تابع ضرر Focal [۲۵] استفاده شده است:

$$L_{prob} = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (۲)$$

در این رابطه، p_t درستی تخمین را نشان می‌دهد و از رابطه زیر محاسبه می‌شود:

$$p_t = \begin{cases} p_{pred} & \text{if } p_{true} = 1 \\ 1 - p_{pred} & \text{otherwise} \end{cases} \quad (۳)$$

که p_{pred} تخمین شبکه برای احتمال وجود گل در این محدوده و p_{true} مقدار صحیح آن است. پارامتر α_t اهمیت هر کدام از دو کلاس را مشخص می‌کند. با تنظیم پارامتر γ می‌توان میزان تاثیر نمونه‌هایی که به خوبی تخمین زده شده‌اند را مشخص کرد. در صورتیکه γ مقداری بزرگتر از ۱ داشته باشد، مقدار ضرر برای نمونه‌هایی که به خوبی تخمین زده شده‌اند و p_t برای آنها نزدیک به ۱ است کاهش می‌یابد تا تمرکز شبکه بر بهبود نمونه‌های دیگر باشد.

ضرر مربوط به تخمین مکان نسبی با L_{loc} نشان داده شده است. به منظور افزایش قوام الگوریتم پیشنهادی نسبت به داده‌های پرت، برای کمی‌سازی این ضرر از تابع نرم ۱ هموار استفاده می‌شود:

$$L_{loc} = L_1^s(dx_{pred} - dx_{true}) + L_1^s(dy_{pred} - dy_{true}) \quad (۴)$$

Faster R-CNN [۲۱] پیشنهاد شد که سرعت و دقت بالاتری را نتیجه داد.

الگوریتم‌هایی که از ایده R-CNN استفاده می‌کنند مبتنی بر دو بخش تولید ناحیه‌های پیشنهادی و دسته‌بندی هستند. رویکرد دیگری که برای آشکارسازی اشیاء وجود دارد استفاده از پنجره لغزان بجای تولید ناحیه‌های پیشنهادی است. در نگاه اول ممکن است تصور شود سرعت الگوریتم‌هایی که از ایده پنجره لغزان استفاده می‌کنند کمتر است. با این وجود، بدلیل همپوشانی بخش عمده‌ای از محاسبات میان پیکسل‌های مجاور، الگوریتم‌هایی که از این ایده استفاده می‌کنند به مراتب سریعتر هستند.

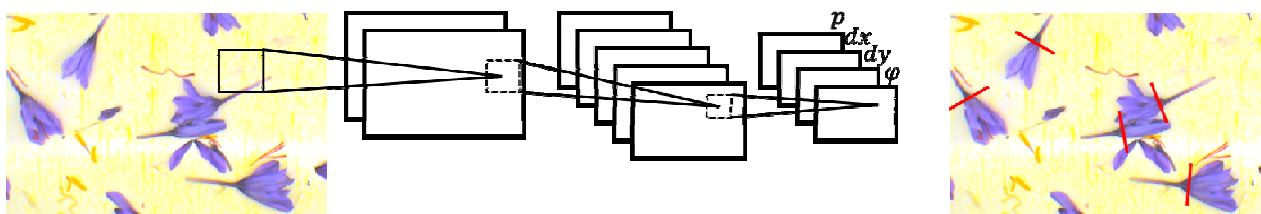
الگوریتم YOLO [۲۲] یکی از نخستین مطالعاتی است که از این رویکرد استفاده می‌کند. در الگوریتم YOLO، تصویر ورودی به تعدادی ناحیه تقسیم می‌شود و برای هر ناحیه، وجود اشیاء و محدوده آن‌ها آموزش داده می‌شود. سپس، الگوریتم SSD [۲۳] با استفاده از یک ساختار چند مقیاسه توانست علاوه بر افزایش دقت، سرعت قابل قبولی را نیز نتیجه دهد. یکی از چالش‌های موجود برای آموزش شبکه‌هایی که از ایده پنجره لغزان استفاده می‌کنند عدم توازن است. به عبارت دیگر، تعداد نمونه‌هایی که شامل اشیاء مورد نظر نیستند به مراتب بیش از سایر نمونه‌ها است. در [۲۴] تابع هزینه‌ای برای CNN پیشنهاد شده است که بتواند در شرایط عدم توازن عملکرد بهتری داشته باشد.

۳ روش پیشنهادی

در این مقاله یک روش مبتنی بر پنجره لغزان برای تشخیص مکان و زاویه گل‌های زعفران پیشنهاد شده است که به صورت نمادین در شکل ۲ نشان داده شده است. در این روش، تصویر ورودی توسط چندین لایه همگشتی پردازش می‌شود تا اطلاعات موجود در همسایگی هر پیکسل استخراج شوند. سپس، با استفاده از لایه Max Pooling ابعاد تصویر کاهش می‌یابد تا حجم محاسبات در لایه‌های بعد کاهش بیابد. در ادامه، چند بار دیگر این ساختار روی خروجی لایه قبل اعمال می‌شود (لایه‌های همگشتی همراه با لایه Max Pooling) تا اطلاعات کامل‌تری با لحاظ کردن همسایگی بزرگتری از تصویر استخراج شوند. خروجی این بخش از شبکه به ازای یک تصویر با ابعاد $W \times H \times 3$ دارای ابعاد $F \times H/2^b \times W/2^b$ خواهد بود که b تعداد لایه‌های Pooling با گام ۲ و F تعداد ویژگی‌های محاسبه شده است. پس از استخراج این ویژگی‌های معنایی از تصویر، برای هر کدام از نقاط چهار پارامتر تخمین زده می‌شود. این پارامترها عبارتند از:

- احتمال وجود نقطه برش گل در این محدوده (p)
- اختلاف نقطه برش نسبت به مرکز محدوده در دو راستای افقی و عمودی (dx, dy)
- زاویه برش گل (φ)

¹ Receptive Field



شکل ۲: نمای شماتیک الگوریتم پیشنهادی برای تشخیص مکان و زاویه گل‌های زعفران

آموزش می‌بیند. در این حالت، بجای تخمین زاویه مطلق گل، تنها نیاز است تا زاویه نسبی آن نسبت به مرکز محدوده مورد نظر تخمین زده شود. به عنوان مثال، همانطور که در شکل ۳ نشان داده شده است، زاویه φ به صورت مجموع $\varphi_0 + d\varphi$ بازنویسی می‌شود که φ_0 توسط دسته‌بند N کلاسه تشخیص داده می‌شود و $d\varphi$ توسط بخش تخمین نسبی زاویه تخمین زده می‌شود (محدوده تغییرات $d\varphi$ بازه $-180/N$ تا $180/N$ است).

به عبارت دیگر، در این روش برای هر محدوده $4N$ پارامتر تعریف می‌شود. در این معماری، زاویه گل به صورت نسبی تخمین زده می‌شود و به همین دلیل دارای دقت بسیار بالاتری است. به عبارت بهتر، ابتدا یک دسته‌بند زاویه گل را از بین N قسمت انتخاب می‌کند و سپس زاویه نسبی که محدوده کوچکتری دارد تخمین زده می‌شود. در بخش بعد نتایج این تقسیم‌بندی بر عملکرد الگوریتم پیشنهادی تحلیل می‌شود.

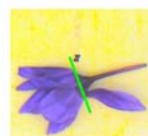
۴ نتایج آزمایش‌ها و بحث

در این بخش ابتدا پایگاه داده مورد استفاده برای ارزیابی الگوریتم پیشنهادی تشریح می‌شود. سپس، معیارهای مورد استفاده برای ارزیابی توضیح داده می‌شوند. در ادامه جزئیات پیاده‌سازی ذکر شده و نتایج بدست آمده گزارش و تحلیل می‌شوند.

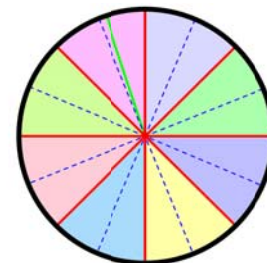
۴-۱ پایگاه داده

به منظور ارزیابی الگوریتم پیشنهادی، یک پایگاه داده شامل ۱۶۳ تصویر با ابعاد 1200×900 جمع‌آوری شده است که در مجموع شامل ۳۰۳۵ گل زعفران هستند و نمونه‌هایی از آنها در شکل ۴ نشان داده شده است. این تصاویر در حالی ثبت شده‌اند که نور نقاله با سرعت 10 cm/s در حال حرکت بوده است. برای نورپردازی، از یک رشته LED در یک محفظه بسته استفاده شده است تا نور محیط تا حد امکان یکنواخت و نامتغیر باشد.

برای هر گل زعفران، مکان و زاویه مناسب برای برش به صورت دستی توسط یک خبره مشخص شده‌اند. به منظور نمایش بر روی تصویر، برای هر گل زعفران پاره‌خطی ترسیم می‌شود که مرکز آن بر روی مکان برش و زاویه آن متناسب با زاویه برش باشد. چند نمونه گل برجسب خورده در شکل ۵ با پاره‌خط‌های قرمز رنگ مشخص شده‌اند. تعداد گل‌های موجود در تصاویر به طور میانگین حدود ۱۹ عدد است. در شکل ۶ هیستوگرام تعداد گل‌های موجود در تصاویر پایگاه داده ترسیم شده است.



(ب)



(الف)

$$\varphi = 109 = 112.5 - 3.5 = \varphi_0 + d\varphi$$

شکل ۳: الف) تقسیم‌بندی ۳۶۰ درجه به ۸ قسمت، و ب) نحوه تبدیل

زاویه یک گل از مقدار مطلق به مقدار نسبی

$$L_1^s(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (5)$$

عبارت سوم از تابع ضرر (۱) مربوط به خطای تخمین زاویه است. از آنجائیکه زاویه دارای رفتار متناوب است، محاسبه فاصله زاویه تخمین زده شده و مقدار هدف معیار مناسبی نیست و از معیار زیر استفاده می‌کنیم:

$$L_{ang} = 1 - \cos(\varphi_{pred} - \varphi_{true}) \quad (6)$$

در این تابع، اختلاف زاویه تخمین زده شده نسبت به زاویه صحیح سنجیده می‌شود و سپس مقدار $\cos$ آن محاسبه می‌شود تا اختلاف زاویه 2π بی‌اثر شود. بنابراین، بهترین تخمین دارای اختلاف زاویه ۰ (یا مضرب صحیحی از 2π) خواهد بود که منجر به ضرر ۰ می‌شود و بدترین تخمین با اختلاف زاویه π (یا مضرب فردی از π) منجر به ضرر ۲ می‌شود. لازم به ذکر است ضررهای مربوط به تخمین مکان و زاویه تنها برای ناحیه‌هایی محاسبه می‌شوند که در آنها گل وجود داشته باشد.

با وجود آنکه تابع ضرر متناسبی برای تخمین زاویه پیشنهاد شده است، نتایج آزمایش‌ها نشان می‌دهند که دقت تخمین زاویه کافی نیست. مهمترین دلیل برای این مشکل پیچیدگی تخمین زاویه گل با استفاده از لایه‌های همگشتی است. به همین دلیل، پیشنهاد می‌شود تا محدوده ۳۶۰ درجه به N قسمت تقسیم شود که در شکل ۳ نحوه تقسیم ۳۶۰ درجه به $N = 8$ قسمت نشان داده شده است.

پس از این تقسیم‌بندی، برای هر قسمت چهار پارامتر ذکر شده تخمین زده شوند. به عبارت دیگر، برای گل‌هایی که زاویه آنها در بازه ۰ تا $360/N$ قرار دارد یک دسته‌بند مجزا آموزش می‌بیند تا پارامترهای مربوط به چنین گل‌هایی را تشخیص دهد. به همین ترتیب، برای هر بازه $360/N$ درجه یک دسته‌بند مجزا



شکل ۴: نمونه تصاویری از پایگاه داده تهیه شده

۲-۴ معیارهای ارزیابی

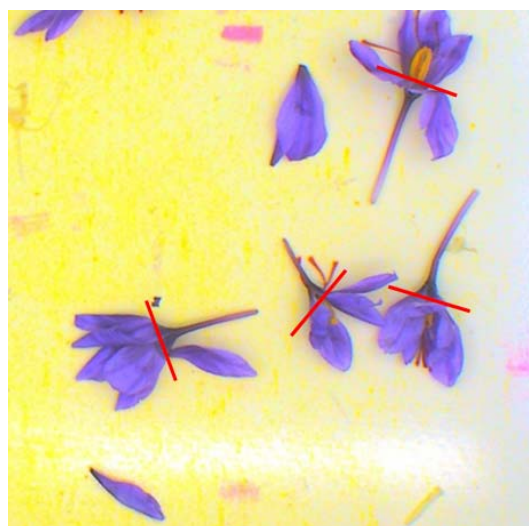
مقدار ضرر نهایی حاصل از الگوریتم پیشنهادی نشان دهنده میزان توانایی شبکه در یادگیری مسأله مورد نظر است. علاوه بر این معیار، از معیار F_1 با رابطه زیر نیز برای ارزیابی عملکرد الگوریتم پیشنهادی استفاده می‌کنیم:

$$F_1 = 2 \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (۷)$$

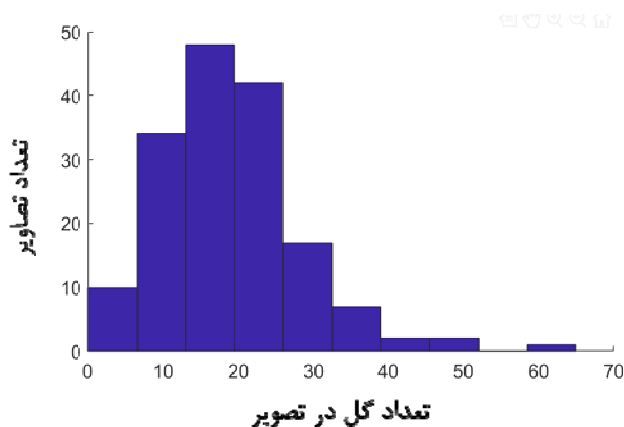
$$\text{recall} = \frac{TP}{TP + FN} \quad (۸)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (۹)$$

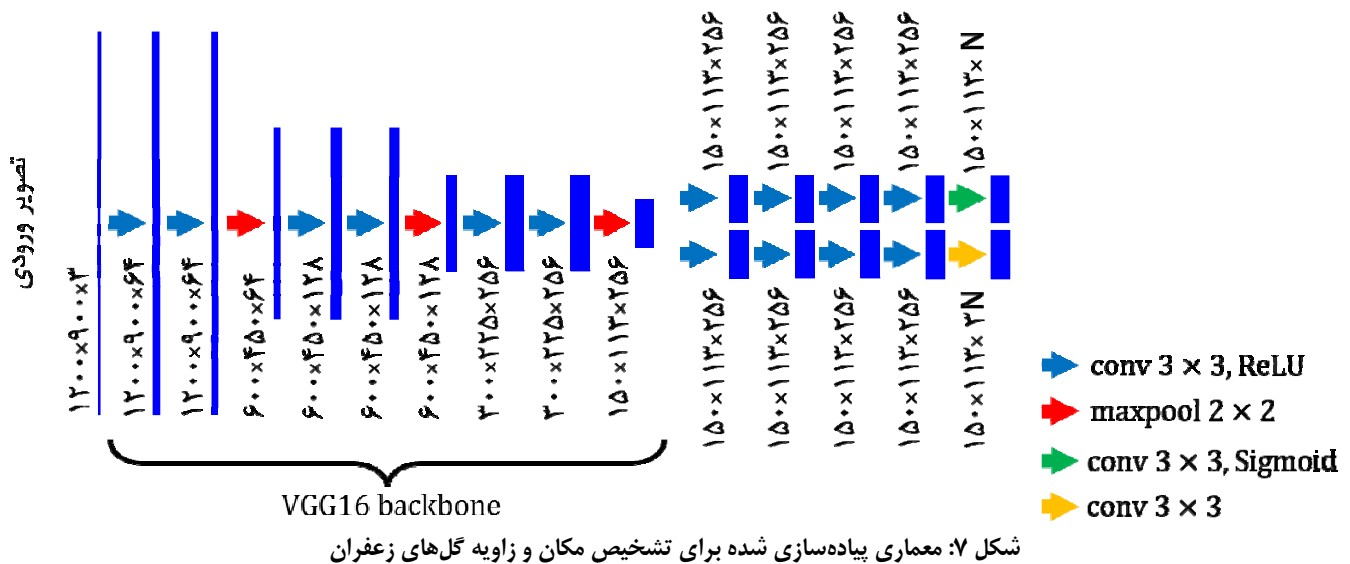
در رابطه‌های قبل، TP نشان دهنده تعداد گل‌هایی است که مکان و زاویه آنها به درستی توسط الگوریتم پیشنهادی تخمین زده شده است. FN برابر با تعداد گل‌هایی است که پارامترهای مربوط به آن به درستی تشخیص داده نشده است. FP برابر با تعداد اشیائی است که توسط الگوریتم پیشنهادی به عنوان پارامترهای گل تخمین زده شده است اما با هیچکدام از گل‌های موجود در تصویر منطبق نشده است. پارامتر صحت (recall) مشخص می‌کند چند درصد از گل‌ها به درستی تشخیص داده شده‌اند و پارامتر دقت (precision) مشخص می‌کند چند درصد از اشیاء تشخیص داده شده به درستی گل بوده‌اند. برای داشتن یک سامانه قابل قبول نیاز است تا هر دوی این معیارها مقدار قابل قبولی داشته باشند که ترکیب آنها در معیار F_1 محاسبه می‌شود. به عبارت دیگر، هرچه مقدار F_1 به ۱ نزدیک‌تر باشد نشان‌دهنده عملکرد بهتر الگوریتم است.



شکل ۵: محل و زاویه برش گل‌ها برای بخشی از تصاویر



شکل ۶: هیستوگرام تعداد گل‌های موجود در تصاویر پایگاه داده



داده شده در الگوریتم ۱ الگوریتم ۱ انجام می شود. در این الگوریتم، ابتدا تصویر ورودی با ابعاد 1200×900 وارد شبکه شده و نقشه های 150×113 محاسبه می شوند. خروجی نخست شبکه N نقشه برای تخمین احتمال وجود گل است. برای هر نقطه ای که احتمال وجود گل بیش از ۰٫۵ باشد، پارامترهای گل با استفاده از نقشه های dx ، dy و $d\phi$ محاسبه می شود. برای مختصات مرکز گل، پس از افزودن اختلاف نسبی به موقعیت نقطه، ضرب ۸ ضرب می شود تا اثر کاهش ابعاد مکانی ویژگی ها در اثر استفاده از سه لایه Pooling جبران شود. برای زاویه گل نیز ابتدا زاویه مطلق گل از k محاسبه شده و سپس اختلاف نسبی به آن اضافه می شود.

استفاده از پنجره لغزان با گام ۸ و تشخیص های جداگانه به ازای زاویه های مختلف می تواند منجر به تشخیص های متعدد یک گل واحد شود. به همین دلیل، پس از استخراج گل ها از تصویر، از الگوریتم حذف مقادیر غیر بیشینه^۲ (NMS) استفاده می شود. در این الگوریتم، ابتدا گلی که دارای بالاترین احتمال باشد انتخاب می شود و تمام گل هایی که با آن مشابه باشند (فاصله مرکز آنها کمتر از ۵ پیکسل و اختلاف زاویه آنها کمتر از ۲۰ درجه باشد) حذف می شوند. سپس، از میان گل های باقی مانده، گلی که دارای بالاترین احتمال باشد انتخاب شده و همانند قبل، گل های مشابه با آن حذف می شوند. این عملیات تا جایی ادامه می یابد که تمام گل ها انتخاب یا حذف شده باشند.

برای پیاده سازی الگوریتم پیشنهادی از چارچوب Keras استفاده شده است. همچنین، برای تقویت داده از چرخش تصادفی (۳۶۰ درجه)، flip، جابجایی، تغییر مقیاس و همچنین افزودن نویز استفاده شده است. سخت افزار مورد استفاده دارای یک کارت گرافیکی GeForce GTX 1050 با حافظه 4G بوده است. برای آموزش پارامترهای شبکه از بهینه ساز Adam با نرخ یادگیری 0.0001 استفاده شده است.

۳-۴ جزئیات پیاده سازی

با توجه به اینکه حجم پایگاه داده به نسبت کوچک است، برای ارزیابی الگوریتم پیشنهادی از روش ارزیابی متقابل با ۵ فولد استفاده می کنیم. در این روش ارزیابی، تصاویر به ۵ گروه تقسیم می شوند. سپس، در هر آزمایش از تصاویر ۴ گروه برای آموزش شبکه و از تصاویر گروه دیگر برای ارزیابی شبکه استفاده می شود. پس از انجام ۵ آزمایش، نتایج ارزیابی آنها با همدیگر تلفیق می شوند تا عملکرد الگوریتم پیشنهادی به درستی سنجیده شود. در تلفیق نتایج، مقادیر TP ، FN و FP برای ۵ آزمایش با یکدیگر جمع می شوند و مقادیر $precision$ ، $recall$ و F_1 با استفاده از این مقادیر محاسبه می شوند.

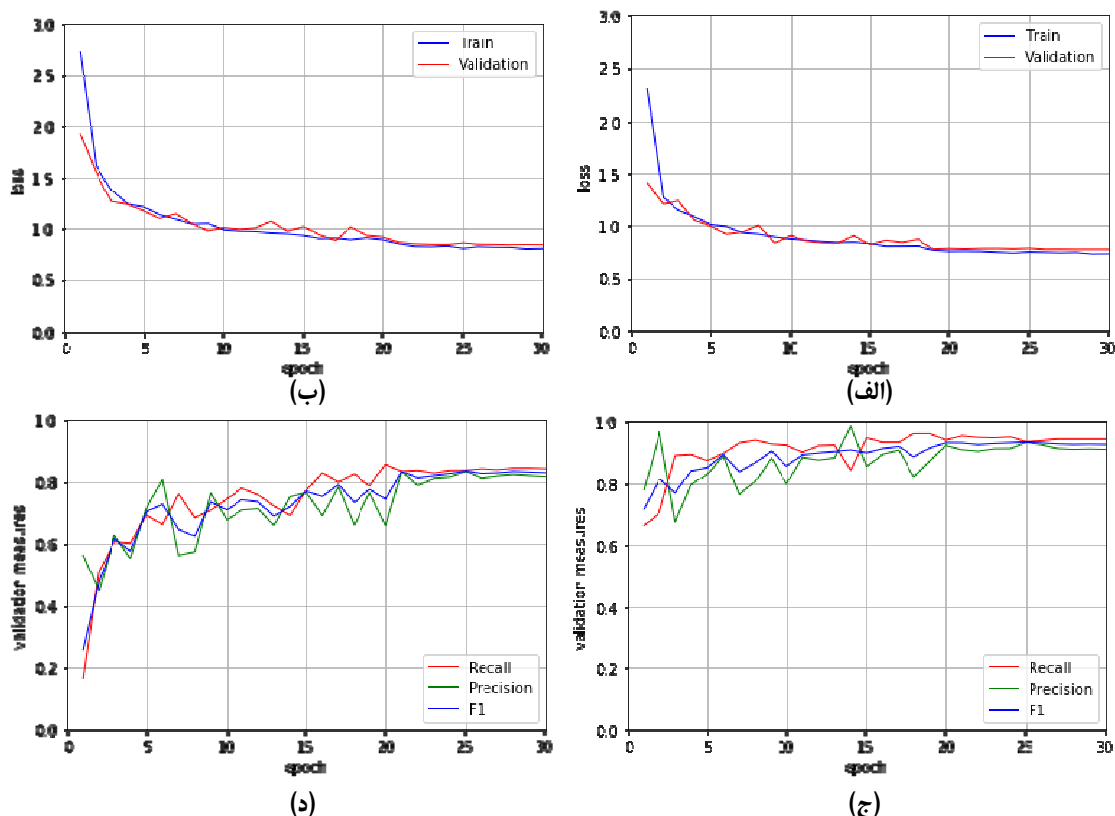
در پیاده سازی انجام شده، از ۳ بلوک ابتدایی شبکه VGG16 [۶] که شامل ۷ لایه همگشتی و ۳ لایه Max Pooling است به عنوان بخش استخراج نقشه ویژگی استفاده شده است. سپس، از ۵ لایه همگشتی برای تخمین احتمال وجود گل و از ۵ لایه همگشتی موازی برای تخمین پارامترهای گل استفاده شده است. این معماری در شکل ۷ ترسیم شده است. با استفاده از این معماری، ناحیه موثر برای محاسبه خروجی در هر نقطه شامل 118×118 پیکسل است که با گام ۸ پیکسل بر روی تصویر ورودی می لغزد. نتایج آزمایش ها نشان می دهد که این معماری می تواند به خوبی گل های نزدیک به یکدیگر را تشخیص بدهد.

به دلیل کوچک بودن پایگاه داده و افزایش سرعت همگرایی از تکنیک یادگیری انتقالی^۱ استفاده شده است. در این روش، وزن های لایه هایی که از معماری VGG کپی شده اند، برابر با وزن هایی قرار داده می شوند که از آموزش بر روی پایگاه داده ImageNet [۲۶] بدست آمده اند.

پس از آموزش مدل بر روی داده های آموزشی، تشخیص پارامترهای گل های زعفران در تصاویر آزمون طبق شبه کد نشان

² Non-Maximum Suppression

¹ Transfer learning



شکل ۸: نمودار تابع ضرر و دقت شبکه پیشنهادی برای فولد نخست، (الف) تابع ضرر بدون تقسیم‌بندی زاویه، (ب) تابع ضرر با تقسیم‌بندی زاویه به ۱۲ قسمت ۳۰ درجه، (ج) معیارهای دقت بدون تقسیم‌بندی زاویه، و (د) معیارهای دقت با تقسیم‌بندی زاویه به ۱۲ قسمت ۳۰ درجه

صفر نمی‌رسد زیرا در مشخص کردن پارامترهای گل‌ها توسط خبره مقداری خطا وجود داشته است. به عبارت دیگر، حتی اگر یک خبره یک تصویر را دو مرتبه حاشیه‌نویسی کند ممکن است زاویه هر گل در حد چند درجه و مکان آن در حد چند پیکسل تغییر کند و طبیعی است که شبکه نتواند مقدار دقیق این پارامترها را تخمین بزند مگر آنکه بیش‌برازش شده باشد. استفاده از روش‌های داده‌افزایی مناسب تاثیر زیادی در عدم بیش‌برازش مدل پیشنهادی داشته است.

نتایج کمی ارزیابی الگوریتم پیشنهادی برای فولدهای پنج‌گانه و مجموع در جدول ۱ گزارش شده است. همانطور که مشاهده می‌شود، الگوریتم پیشنهادی توانسته است در مجموع پارامترهای بیش از ۹۵٪ از گل‌ها را به درستی تخمین بزند. همچنین، دقت الگوریتم پیشنهادی در حدود ۹۲٪ است.

جدول ۱: عملکرد الگوریتم پیشنهادی

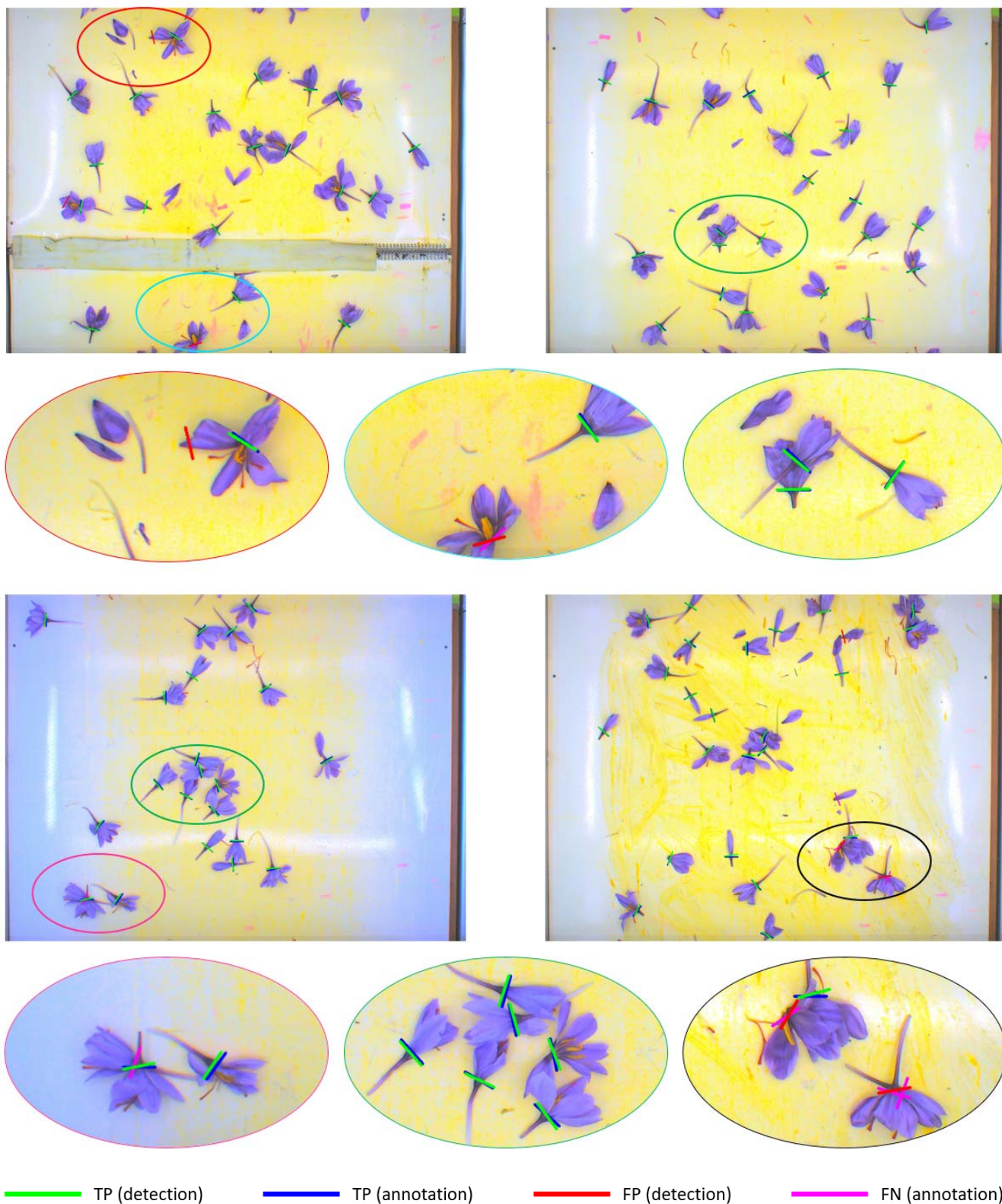
F1	دقت	صحت	FP	FN	TP	
۹۲,۸۰	۹۱,۱۶	۹۴,۵۰	۵۵	۳۳	۵۶۷	فولد ۱
۹۳,۲۳	۹۱,۲۸	۹۵,۲۷	۵۲	۲۷	۵۴۴	فولد ۲
۹۴,۱۴	۹۲,۶۲	۹۵,۷۱	۴۸	۲۷	۶۰۲	فولد ۳
۹۴,۴۹	۹۲,۵۵	۹۶,۵۰	۵۱	۲۳	۶۳۴	فولد ۴
۹۲,۸۰	۹۲,۰۱	۹۳,۶۰	۴۷	۳۷	۵۴۱	فولد ۵
۹۳,۵۲	۹۱,۹۵	۹۵,۱۶	۲۵۳	۱۴۷	۲۸۸۸	مجموع

الگوریتم ۱: نحوه استخراج پارامترهای گل‌ها در فاز آزمون

- اعمال شبکه شکل ۷ بر تصویر ورودی
- برای هر نقطه (i, j, k) در نقشه‌های احتمال به شرط $p(i, j, k) \geq 0.5$
 - o $x = 8 \times (i + dx(i, j, k))$
 - o $y = 8 \times (j + dy(i, j, k))$
 - o $\varphi = \frac{2\pi}{N} \times (k + 0.5) + d\varphi(i, j, k)$
 - o افزودن (x, y, φ) به گل‌های تشخیص داده شده
- حذف مقادیر غیربیشینه

۴-۴ نتایج

نمودار تابع ضرر بر حسب دوره آموزش در شکل ۸ برای فولد نخست ترسیم شده است. در این شکل، قسمت (الف) مربوط به حالتی است که زاویه گل به صورت مطلق سنجیده می‌شود و تقسیم‌بندی زاویه انجام نمی‌شود. قسمت (ب) نیز مربوط به حالتی است که بازه ۳۶۰ درجه به ۱۲ قسمت ۳۰ درجه تقسیم شده است و تخمین زاویه به صورت نسبی انجام می‌شود. همانطور که مشاهده می‌شود، مقدار تابع ضرر برای هر دو بخش آموزش و ارزیابی به خوبی کاهش می‌یابد که نشان می‌دهد یادگیری شبکه به خوبی انجام می‌شود. البته مقدار ضرر برای داده‌های آموزشی هم به



شکل ۹: نمونه‌های از عملکرد الگوریتم پیشنهادی برای تخمین مکان و زاویه برش گل‌ها

(FP) با رنگ قرمز ترسیم شده‌اند. پاره‌خط‌هایی که توسط خبره مشخص شده‌اند نیز در صورتیکه با یکی از خروجی‌های الگوریتم منطبق باشند (TP) با رنگ آبی و در غیر این صورت (FN) با رنگ بنفش ترسیم شده‌اند.

چهار نمونه تصویر از نتایج الگوریتم پیشنهادی در شکل ۹ نشان داده شده است. در این شکل، پاره‌خط‌هایی که توسط الگوریتم پیشنهادی تخمین زده شده‌اند با دو رنگ سبز و قرمز نشان داده شده‌اند. رنگ سبز مربوط به تخمین‌هایی است که با یکی از گل‌های برجسب خورده منطبق شده‌اند (TP) و مابقی (

جدول ۲: مقایسه الگوریتم پیشنهادی به ازای تقسیم‌بندی‌های مختلف زاویه

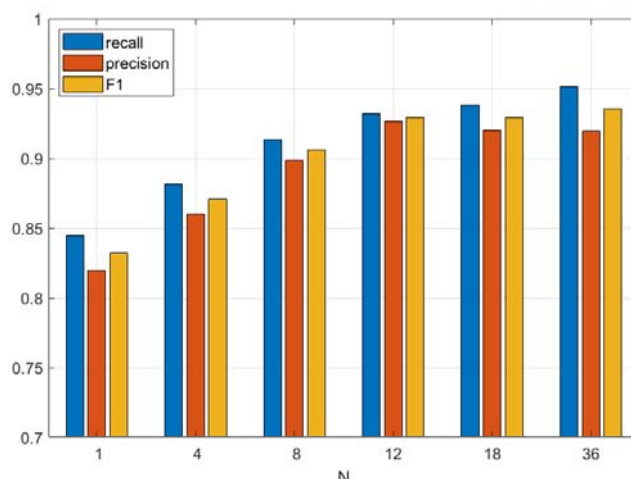
تعداد تقسیم	زاویه هر قسمت	صحت	دقت	F1	تعداد پارامترهای شبکه	زمان متوسط هر تصویر (ثانیه)
۱	۳۶۰	۸۴,۴۴	۸۱,۹۸	۸۳,۲۰	۷۱۲۱۲۲۰	۰,۳۵۲
۴	۹۰	۸۸,۱۷	۸۶,۰۲	۸۷,۰۸	۷۱۴۸۸۸۰	۰,۳۶۱
۸	۴۵	۹۱,۳۳	۸۹,۸۴	۹۰,۵۸	۷۱۸۵۷۶۰	۰,۳۷۶
۱۲	۳۰	۹۳,۲۱	۹۲,۶۳	۹۲,۹۲	۷۲۲۲۶۴۰	۰,۳۹۰
۱۸	۲۰	۹۳,۸۵	۹۲,۰۴	۹۲,۹۴	۷۲۷۷۹۶۰	۰,۴۰۱
۳۶	۱۰	۹۵,۱۶	۹۱,۹۵	۹۳,۵۲	۷۴۴۳۹۲۰	۰,۴۶۹

همپوشانی گل‌ها و زمینه پیچیده نمی‌توانند کارآیی لازم را داشته باشند.

به منظور مقایسه با روش‌های سنتی پردازش تصویر، یک الگوریتم مبتنی بر تطبیق کلیشه^۱ پیاده‌سازی کرده‌ایم که جزئیات آن در ادامه مطرح می‌شود. در فاز آموزش، تمام گل‌های موجود در تصاویر آموزشی به صورت عمودی و با عرض ثابتی حول مرکز برش زده می‌شوند و با استفاده از روش خوشه‌بندی سلسله‌مراتبی^۲ به چند گروه تقسیم می‌شوند. سپس، میانگین تصاویر هر خوشه به عنوان یک کلیشه در نظر گرفته می‌شود. در شکل ۱۱ چند نمونه از کلیشه‌های آموخته شده برای تشخیص گل زعفران در تصویر نشان داده شده است.

در فاز آزمون، تطبیق کلیشه میان تمام کلیشه‌ها با تصویر ورودی محاسبه می‌شود و نقاطی که پس از عملیات حذف مقادیر غیربیشینه دارای تطبیق بزرگی باشند، به عنوان محل وجود گل زعفران اعلام می‌شوند. علاوه بر این، به منظور تشخیص گل‌هایی که دارای زاویه‌های متفاوتی هستند، هر کلیشه در زاویه‌های متفاوتی چرخانده می‌شود و به طور جداگانه، انطباق آنها با تصویر محاسبه می‌شود.

نتایج این روش به ازای تعداد کلیشه‌های مختلف و تعداد زاویه‌های مختلف در مقایسه با روش پیشنهادی مبتنی بر یادگیری عمیق در جدول ۳ گزارش شده است. همانطور که در این جدول قابل مشاهده است، با افزایش تعداد کلیشه‌های مورد استفاده و همچنین با افزایش تعداد زاویه‌هایی که تطبیق کلیشه انجام می‌شود، تعداد گل‌های بیشتری به درستی تشخیص داده می‌شوند و معیار صحت افزایش می‌یابد. در سوی مقابل، زمان متوسط پردازش برای هر تصویر نیز افزایش می‌یابد. همانطور که مشاهده می‌شود، با وجود پیاده‌سازی بر روی GPU، باز هم سرعت روش مبتنی بر تطبیق کلیشه برای دستیابی به دقت مناسب بیش از روش مبتنی بر یادگیری عمیق است در حالیکه دقت روش مبتنی بر یادگیری عمیق به مراتب از روش مبتنی بر تطبیق کلیشه بالاتر است. این مقایسه نشان می‌دهد که استفاده مناسب از روش‌های مبتنی بر یادگیری



شکل ۱۰: عملکرد الگوریتم پیشنهادی به ازای تقسیم‌بندی‌های مختلف

همانطور که در شکل ۹ مشاهده می‌شود، بخش قابل توجهی از گل‌ها به درستی تشخیص داده شده‌اند و تعداد زیادی از FP مربوط به گلبرگ‌های گل زعفران است. همچنین، در برخی از موارد وضعیت گل‌ها پیچیده است و به همین دلیل مکان و جهت گل با خطا تخمین زده شده است.

به منظور مقایسه اثر تقسیم‌بندی و تخمین نسبی زاویه، معیارهای $recall$ ، $precision$ و F_1 بر حسب تعداد قسمت‌های تقسیم‌بندی زاویه (N) در شکل ۱۰ ترسیم شده‌اند. همانطور که قابل مشاهده است، با تقسیم‌بندی زاویه گل‌ها دقت و صحت الگوریتم پیشنهادی به اندازه قابل توجهی بهبود می‌یابد. بهبود بدست آمده با تقسیم‌بندی زاویه منجر به شبکه‌ای با پارامترهای بیشتر و زمان محاسباتی بیشتر می‌شود که مقایسه این پارامترها در جدول ۲ انجام شده است.

۴-۵ مقایسه با روش‌های سنت پردازش تصویر

همانطور که در قسمت‌های قبل ذکر شد، الگوریتم‌هایی برای تشخیص مکان و زاویه گل زعفران در تصویر بدون استفاده از یادگیری عمیق توسعه یافته‌اند. بسیاری از روش‌های موجود مبتنی بر دوسطحی سازی تصویر هستند که در شرایط کاربردی با وجود

¹ Template Matching

² Hierarchical Clustering

عمیق می تواند نسبت به روش های سنتی پردازش تصویر برای حل مسائل پیچیده، کارایی بسیار بالاتری را نتیجه دهد.

جدول ۳: مقایسه الگوریتم پیشنهادی با روش مبتنی بر تطبیق کلیشه

زمان متوسط هر تصویر (ثانیه)	F1	دقت	صحت	FP	FN	TP	تعداد زاویه	تعداد کلیشه	
۰٫۱۹۴	۵۰٫۵۹	۶۱٫۲۴	۴۳٫۱۰	۸۲۸	۱۷۲۷	۱۳۰۸	۱۲	۵	روش مبتنی بر تطبیق کلیشه
۰٫۲۸۶	۵۴٫۴۷	۶۰٫۴۷	۴۹٫۵۶	۹۸۳	۱۵۳۱	۱۵۰۴	۱۸	۵	
۰٫۵۶۴	۵۶٫۴۲	۵۷٫۷۴	۵۵٫۱۶	۱۲۲۵	۱۳۶۱	۱۶۷۴	۳۶	۵	
۱٫۱۱۲	۵۹٫۳۷	۵۸٫۵۳	۶۰٫۲۳	۱۲۹۵	۱۲۰۷	۱۸۲۸	۳۶	۱۰	
۲٫۲۶۴	۶۲٫۰۳	۵۸٫۶۰	۶۵٫۹	۱۴۱۳	۱۰۳۵	۲۰۰۰	۳۶	۲۰	
۰٫۴۶۹	۹۳٫۵۲	۹۱٫۹۵	۹۵٫۱۶	۲۵۳	۱۴۷	۲۸۸۸	روش پیشنهادی مبتنی بر یادگیری عمیق		

مراجع

- [1] <https://en.irna.ir/news/83001760/Official-Iran-accounts-for-92-of-world-saffron-output>. 2018.
- [2] خلیلی، خلیل و سراج پور، مسعود، "اتوماسیون برش گل زعفران با استفاده از پردازش تصویر"، چهارمین کنفرانس ماشین بینایی و پردازش تصویر، ۱۳۸۵.
- [3] بخشی، حامد، گلزاریان، محمود رضا و عباسپور، محمد حسین، "الگوریتم بینایی رایانه ای برای ردیف کردن گل زعفران"، هشتمین کنگره ملی مهندسی ماشین های کشاورزی (بیوسیستم) و مکانیزاسیون ایران، ۱۳۹۲.
- [4] علی آبادی، رضا، محمدی، مجید و بیاتی، عاطفه، "بهبود روش های اتوماسیون برش گل زعفران با استفاده از تکنیک های هوشمند"، اولین کنفرانس ملی مهندسی برق اصفهان، ۱۳۹۱.
- [5] http://www.fartaktrade.com/product_en.php?id=17. 2020.
- [6] Simonyan, Karen and Zisserman, Andrew, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.
- [7] Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, and Rabinovich, Andrew, "Going deeper with convolutions", Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
- [8] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
- [9] Girshick, Ross, Donahue, Jeff, Darrell, Trevor, and Malik, Jitendra, "Region-based convolutional networks for accurate object detection and segmentation", IEEE transactions on pattern analysis machine intelligence, Vol. 38, pp. 142-158, 2016.
- [10] Mohammadi, Mohammad Reza, "Deep Multiple Instance Learning for Airplane Detection in High



شکل ۱۱: نمونه هایی از کلیشه های آموخته شده

۵ جمع بندی

در این مقاله، یک الگوریتم جدید برای تشخیص مکان و زاویه برش گل های زعفران بر اساس شبکه های همگشتی عمیق پیشنهاد شد. در الگوریتم پیشنهادی، یک شبکه همگشتی به صورت پنجره لغزان با گام ۸ (به دلیل استفاده از ۳ لایه Pooling) بر روی تصویر لغزانده می شود. اگر تصویر ورودی دارای ابعاد $W \times H \times 3$ باشد، خروجی شبکه دارای ابعاد $W/8 \times H/8 \times 4N$ است. در هر نقطه، احتمال وجود گل، زاویه نسبی و مختصات نسبی توسط شبکه تخمین زده می شوند. به دلیل قابلیت پیاده سازی موازی بر روی GPU، این پنجره لغزان دارای سرعت قابل قبولی است.

نتایج پیاده سازی الگوریتم پیشنهادی بر روی یک پایگاه داده جمع آوری شده نشان می دهد که بیش از ۹۵٪ از گل ها به درستی تخمین زده می شوند. الگوریتم پیشنهادی بر روی یک دستگاه نیمه صنعتی همراه با قسمت های مکانیکی پیاده سازی شده است و نتایج عملی قابل قبول گزارش شده است. در این سیستم، ابتدا گل ها بر روی نوار نقاله ریخته می شوند. سپس، از نوار نقاله تصویربرداری شده و با استفاده از الگوریتم پیشنهادی در این مقاله، مکان و زاویه گل های زعفران تخمین زده می شود. سپس، فرمان لازم به یک بازوی مکانیکی داده می شود تا برش های مورد نظر را انجام دهد (کالیبراسیون مربوط به موقعیت نسبی دوربین و بازو نیز به صورت آفلاین انجام شده است). در نهایت، در بخش جداساز، کلاله های گل زعفران از باقی اجزاء جدا می شوند.

detector", European conference on computer vision, 2016.

- [24] Lin, Tsung-Yi, Goyal, Priya, Girshick, Ross, He, Kaiming, and Dollár, Piotr, "Focal loss for dense object detection", Proceedings of the IEEE international conference on computer vision, 2017.
- [25] Lin, Tsung-Yi, Goyal, Priyal, Girshick, Ross, He, Kaiming, and Dollár, Piotr, "Focal loss for dense object detection", IEEE transactions on pattern analysis machine intelligence, 2018.
- [26] Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, and Fei-Fei, Li, "Imagenet: A large-scale hierarchical image database", Proceedings of the IEEE conference on computer vision and pattern recognition, 2009.



محمدرضا محمدی دانش‌آموخته دکترای مهندسی برق از دانشگاه صنعتی شریف در سال ۱۳۹۴ است. وی از سال ۱۳۹۶ استادیار گروه هوش مصنوعی در دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران است. زمینه‌های تحقیقاتی مورد علاقه وی شامل بینایی ماشین، پردازش تصویر و ویدئو، یادگیری ماشین و یادگیری عمیق است و توانسته مقاله‌های متعددی را در این زمینه‌ها به چاپ برساند و چندین پروژه صنعتی را در حوزه هوش مصنوعی انجام دهد.



مهرداد کرمی دوره کارشناسی مهندسی مکانیک را در دانشگاه صنعتی امیرکبیر و دوره کارشناسی ارشد مهندسی مکترونیک را در دانشگاه صنعتی شریف گذرانده است. وی در سال ۱۳۹۶ برای طرح دستگاه جداسازی اجزای گل زعفران، مبتنی بر پردازش تصویر، توانست رتبه ۳ جشنواره جوان خوارزمی را بدست بیاورد. وی در حال حاضر مدیر عامل شرکت سامانه‌های هوشمند کشاورزی والی است.

Resolution Imagery", arXiv preprint arXiv:1808.06178, 2018.

- [11] Chen, Liang-Chieh, Papandreou, George, Kokkinos, Iasonas, Murphy, Kevin, and Yuille, Alan, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs" IEEE transactions on pattern analysis machine intelligence, Vol. 40, pp. 834-848, 2018.
- [12] Noh, Hyeonwoo, Hong, Seunghoon, and Han, Bohyung, "Learning deconvolution network for semantic segmentation", Proceedings of the IEEE international conference on computer vision, 2015.
- [13] Lai, Hanjiang, Xiao, Shengtao, Pan, Yan, Cui, Zhen, Feng, Jiashi, Xu, Chunyan, Yin, Jian, and Yan, Shuicheng, "Deep recurrent regression for facial landmark detection", IEEE Transactions on Circuits Systems for Video Technology, Vol. 28, pp. 1144-1157, 2018.
- [14] Sun, Yi, Wang, Xiaogang, and Tang, Xiaoou, "Deep convolutional network cascade for facial point detection", Proceedings of the IEEE conference on computer vision and pattern recognition, 2013.
- [15] Zhang, Zhanpeng, Luo, Ping, Loy, Chen Change, and Tang, Xiaoou, "Facial landmark detection by deep multi-task learning", European Conference on Computer Vision, 2014.
- [16] Mureşan, Horea, and Oltean, Mihai, "Informatica, Fruit recognition from images using deep learning", Vol. 10, pp. 26-42, 2018.
- [17] Sa, Inkyu, Ge, Zongyuan, Dayoub, Feras, Upcroft, Ben, Perez, Tristan, and McCool, Chris, "Deepfruits: A fruit detection system using deep neural networks", Sensors, Vol. 16, pp. 1222, 2016.
- [18] Bargoti, Suchet, and Underwood, James, "Deep fruit detection in orchards", IEEE International Conference on Robotics and Automation (ICRA), 2017.
- [19] Uijlings, Jasper RR, Van De Sande, Koen EA, Gevers, Theo, and Smeulders, Arnold WM, "Selective search for object recognition", International journal of computer vision, Vol. 104, pp. 154-171, 2013.
- [20] Girshick, Ross, "Fast r-cnn", Proceedings of the IEEE international conference on computer vision, 2015.
- [21] Ren, Shaoqing, He, Kaiming, Girshick, Ross, and Sun, Jian, "Faster r-cnn: Towards real-time object detection with region proposal networks", Advances in neural information processing systems, 2015.
- [22] Redmon, Joseph, Divvala, Santosh, Girshick, Ross, and Farhadi, Ali, "You only look once: Unified, real-time object detection", Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
- [23] Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, Fu, Cheng-Yang, and Berg, Alexander C, "SSD: Single shot multibox