



طرح سرویس شناسایی اشیاء



شرکت فناوری اطلاعات و ارتباطات پاسارگاد آریان
ICT Holding of Pasargad Financial Group
www.fanap.ir

سند مشخصات طرح سرویس

Object Detection

مورد نیاز شرکت فناپ

هدف از این طرح تدوین مشخصات سرویس مورد نیاز شرکت فناوری اطلاعات و ارتباطات پاسارگاد آریان (فناپ) در حوزه Object Detection بوده که می‌بایست توسط مجری یا مجریان احتمالی طراحی و پیاده‌سازی گردد. در این راستا و در ادامه، خروجی طرح (Deliverables)، شرح نیازمندیها و محدودیت زمانی اجرا طرح مشخص شده است. همچنین به مواردی که می‌بایست در تخمین بودجه مورد نیاز متقاضیان اجرای طرح لحاظ گردد، نیز اشاره شده است.

۱. خروجی مورد نیاز

دریافت سرویسی برای مسئله Low-Shot Generic Object Detection با قابلیت تعیین Pixel Mask و طبقه‌بندی اشیاء موجود در تصاویر ورودی به‌مراه امکان افزودن کلاس‌های جدید اشیاء با استفاده از تصاویر محدود، دارا بودن ویژگی تشخیص اشیاء از زوایای مختلف، داشتن پایداری (Robustness) مطلوب در مقابل تخریب‌های مختلف موجود در تصاویر ورودی و ترجیحاً بهره‌مند بودن از خصیصه شناسایی Ontological. همچنین نیاز است که خروجی طرح، معیارهای ارزیابی در مجموعه دادگان قیده شده در این سند را برآورده نماید.

۲. تشریح نیازمندی طرح

در این بخش ابتدا مشخصات اصلی طرح مطرح شده و سپس نحوه اعتبار سنجی آن مشخص می‌گردد. در انتها نیز، قابلیت‌های فنی مورد نیاز در طرح تشریح شده است.

۱.۲. مشخصات اصلی

با توجه به افزایش روزافزون کاربردهای سرویس‌ها و برنامه‌های هوش مصنوعی در زندگی روزمره انسان‌ها، پیش‌برد ابزارهای پایه برای دستیابی به شناخت از محیط زندگی بشر و Context رفتاری انسان‌ها بسیار پر اهمیت می‌باشد. یکی از مهم‌ترین پایه‌های ترین ابزارهای لازم برای شناخت محیط پیرامونی ما توسط الگوریتم‌های هوشمند و ربات‌ها، Object Detection می‌باشد.

مساله تشخیص اشیاء در حوزه بینایی ماشین بر دو قسم است:

- تشخیص اشیاء خاص: به عنوان مثال تشخیص اینکه شیء‌ای خاص همانند برج ایفل در تصویری مشاهده شده یا خیر.
- تشخیص اشیاء عام: تشخیص رده شیء برای تمامی انواع اشیاء موجود در تصویر. به عنوان مثال گربه، سگ، میز و... و همچنین برای دسته گربه‌ها مثلاً انواع متفاوتی گربه وجود دارد از گربه‌های پرشین گرفته تا Siamese Cat و... همه این انواع گربه باید در دسته گربه قرار گیرند.

موضوع پروژه تعریف شده در این سند مربوط به مسئله تشخیص اشیاء عام (Generic Object Detection) می‌باشد. در دنیایی که زندگی می‌کنیم، اشیاء بسیار زیادی وجود دارند. دانشمندان در سالیان اخیر بیشتر به استفاده و پیگیری اشیاء پرداختند که ساختیافته باشند (همانند ماشین‌ها، چهره‌ها، دوچرخه‌ها، هواپیماها و...) یا چندبخشی باشند (همانند، گاو، اسب و انسان‌ها و...) بنابراین اشیایی بی‌ساختار (همانند، آسمان، دریا، چمن، ابرها و...) مورد توجه نبوده‌اند.

فرایند Object Detection، هم شامل فرایند Localization یا مکان‌یابی اشیاء می‌شود و هم شامل فرایند بازشناسی یا Recognition و Classification اشیاء. خروجی این فرایند Bounding Box و یا Pixel Mask اشیاء یافت شده در تصویر و انتساب یک کلاس و برچسب به هر یک از اشیاء شناسائی شده می‌باشد.

در نتیجه موضوع پروژه اصلی مورد نظر عبارت است از طراحی و پیاده‌سازی یک سرویس Generic Object Detection. به نحوی که بتواند در تصاویر ورودی اشیاء را شناسائی کرده و Pixel Mask و برچسب آن‌ها را تعیین نماید.

۲.۲. نحوه اعتبار سنجی طرح

بمنظور سنجش اعتبار طرح مورد نظر، ابتدا معیارهای ارزیابی و مجموعه دادگان مورد نظر معرفی و سپس مقادیر مورد انتظاری که می‌بایست سرویس پیاده‌سازی شده به آن دست یابد، مشخص گشته است. در واقع، روش ارائه شده می‌تواند از هر رویکرد دلخواه یک-مرحله‌ای (One-Stage) و یا دو مرحله‌ای (Two-Stage) استفاده نماید و تنها کافیست به دقت و مقادیر تعیین شده برای معیارهای ارزیابی در مجموعه دادگان مذکور دست پیدا نماید.

۱.۲.۲. معیارهای ارزیابی

معیارهای ارزیابی در نظر گرفته شده جهت صحت سنجی اعتبار این طرح به قرار ذیل است.

- ارزیابی کیفیت Pixel Mask های استخراج شده برای اشیاء با استفاده از سنجه IoU با فرمول زیر انجام می‌شود:

$$\text{IoU}(b_{\text{pred}}, b_{\text{gt}}) = \frac{\text{Area}(b_{\text{pred}} \cap b_{\text{gt}})}{\text{Area}(b_{\text{pred}} \cup b_{\text{gt}})}$$

در فرمول فوق b_{pred} مربوط به Pixel Mask شیء شناسائی شده و b_{gt} مربوط به Pixel Mask معیار یا (Ground Truth) می‌باشد.

- برای ارزیابی دقت Classification از سنجه Mean Average Precision (mAP) با فرمول زیر استفاده خواهد شد:

$$\text{MAP} = \frac{\sum_{q=1}^Q \text{AveP}(q)}{Q}$$

در فرمول فوق Q تعداد تصاویر مورد بررسی و AveP همان AP یا Average Precision است که با فرمول زیر محاسبه می‌شود:

$$\text{AveP} = \sum_{k=1}^n P(k) \Delta r(k)$$

در فرمول فوق P همان Precision است و $\Delta r(k)$ نماینده تغییرات Recall از $k-1$ تا k می‌باشد.

- به منظور ارزیابی سرعت سرویس ارائه شده از سنجه Frame Per Second (FPS) استفاده خواهد شد. در سرویس ارائه شده می‌بایست FPS حداقل برابر ۲۰ با اجرای مدل بر روی یک GPU مدل nVidia 1080 Ti GTX را ارائه دهد.

۲.۲.۲. مجموعه دادگان

مجموعه دادگان مورد نظر جهت اعمال معیارهای ارزیابی در خروجی طرح، عبارتند از:

- Pascal VOC2012 با ۲۰ کلاس مختلف
- MSCOCO با ۸۰ کلاس مختلف
- Open Images با ۶۰۰ کلاس مختلف
- LVIS با بیش از ۱۰۰۰ کلاس مختلف
- ImageNet با ۲۰۰ کلاس مختلف

۳.۲.۲. تعیین مقادیر معیارها در مجموعه دادگان

در جدول ۱ معیارهای ارزیابی متداول مانند AP , AP_{50} , AP_{75} , AP_S , AP_M , AP_L بر روی برخی از دادگان فوق معرفی شده است. در واقع این معیارها ترکیبی از موارد مطرح شده در قسمت ۱.۲.۲ می باشد.

جدول ۱. معیارهای ارزیابی متداول با برخی از مجموعه دادگان

مجموعه دادگان	معیارهای متداول
PascalVOC2012	MAP at 0.5 IOU threshold over all 20 classes.
OpenImages	MAP at 0.5 IOU threshold over 500 most frequent classes.
MSCOCO	• AP_{COCO} : MAP averaged over ten Ω : {0.5 : 0.05 : 0.95};
	• AP_{50} : MAP at 0.5 IoU threshold;
	• AP_{75} : MAP at 0.75 IoU threshold;
	• AP_S : AP_{COCO} for small objects of area smaller than 32^2 ;
	• AP_M : AP_{COCO} for objects of area between 32^2 and 96^2 ;
	• AP_L : AP_{COCO} for large objects of area bigger than 96^2 ;

در جدول زیر نتایج مورد انتظار برای سرویس تحویلی و بر اساس معیارهای معرفی شده، نمایش داده شده است:

جدول ۲. مقادیر مورد انتظار معیارهای ارزیابی در مجموعه دادگان

مجموعه دادگان	معیارهای ارزیابی	نتیجه مورد انتظار
PascalVOC2012	MAP	85%
MSCOCO	AP_{COCO} , AP_{50} , AP_{75} , AP_S , AP_M , AP_L	$AP_{COCO} = 55\%$
		$AP_{50} = 70\%$
		$AP_{75} = 55\%$
		$AP_S = 35\%$
		$AP_M = 55\%$
		$AP_L = 65\%$
ImageNet	Overall AP	75%
OpenImagesV4	MAP	40%

شایان ذکر است که در جدول فوق تنها برخی از مجموعه دادگان و نسخه‌های خاصی از آن‌ها آورده شده. به عنوان مثال برای دادگان OpenImages نسخه ۵ و ۶ نیز موجود است.

۳.۲. قابلیت های سرویس

سرویس مورد نظر می بایست از قابلیت های ذیل برخوردار باشد.

- تعداد کلاس ها و Category های قابل شناسائی باید اجتماعی از کلاس ها و Category های پنج مجموعه دادگان معرفی شده باشد. همچنین یکی از مهمترین قابلیت های سرویس Object Detection ارائه شده می بایست این باشد که قابلیت افزودن کلاس های جدید اشیاء به مدل با استفاده از تعداد اندکی از تصاویر مربوط به آن کلاس ها، امکان پذیر باشد. تعداد تصاویر این اشیاء برای افزودن آن ها به مدل، کمتر از ۲۰ تصویر خواهد بود. در حقیقت این یک مسئله Low-Shot Object Detection می باشد.
- شناسائی Main Category های اشیاء و درحقیقت شناسائی Ontological اشیاء مؤکداً پیشنهاد می شود. این بدین معنی است که اگر یک شیء دقیقاً قابل تشخیص و انتساب به کلاس خاصی مانند اسب نیست، حداقل کلاس کلی آن مثلاً حیوان تشخیص داده شود.
- مطمئناً تشخیص اشیاء در Viewpoint های مختلف از اهداف اصلی پروژه است. در نتیجه تشخیص اشیاء از منظر دید راست، چپ، پشت و جلو و...، کاملاً باید در نظر گرفته شود.
- سرویس ارائه شده می بایست نسبت به موارد زیر به مقدار زیادی Robust باشد:
 - ✓ تغییرات نور (Illumination)
 - ✓ تغییرات شکلی اشیاء (Deformation)
 - ✓ تغییرات اندازه (Scale and Size)
 - ✓ پوشیده بودن قسمتی از اشیاء (Partially Occlusion)
 - ✓ تار بودن تصویر (Blurriness)
 - ✓ تغییرات حرکتی اشیاء (Motion)
 - ✓ پائین بودن کیفیت تصویر (Low Resolution)

لازم به ذکر است که میزان Robustness نسبت به هر یک از موارد فوق به صورت دقیق در اسناد بعدی تعیین می گردد.

۳. محدودیت زمانی اجرای طرح

این طرح پس از توافق طرفین می بایست حداکثر در یک بازه ی زمانی شش ماهه به اتمام رسد.

۴. تخمین بودجه اجرای طرح

تخمین بودجه اجرای طرح در قالب هزینه های نفر/ساعت نیروی انسانی، سربار و تجهیزات جانبی مورد نیاز می بایست مشخص شود.